HABITLAB:
IN-THE-WILD BEHAVIOR CHANGE EXPERIMENTS AT SCALE

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Geza Kovacs
May 2019

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Michael Bernstein) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(James Landay)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(BJ Fogg)

Approved for the Stanford University Committee on Graduate Studies

Preface

Behavior change systems help people manage their time online. However, existing productivity systems have tended to assume a one-size-fits all solution, whereas there are many factors - novelty effects, attrition, influences from other apps and devices, and differences in individual motivation - that we must take into account. In this thesis we present HabitLab, an in-the-wild experimentation platform we developed for conducting behavior change experiments, as well as a set of studies we run on the platform. HabitLab is a browser extension and android app with over 12,000 daily active, voluntary users, that users install to help them reduce time online and on their phones. It works by displaying one of 20+ interventions whenever they open an app or site they wish to spend less time on.

We use HabitLab as a large-scale experiment platform to understand behavior change. In our first set of studies, we investigate novelty effects of interventions, finding that compared to always showing the same intervention, a strategy of rotating between different interventions improves intervention effectiveness, but at the cost of increased attrition. This attrition is partly due to users being unfamiliar with rotating interventions, and improving users' mental models with a notice shown whenever a new intervention is shown is able to reduce this attrition. In our second set of studies, we investigate whether reducing time on one site or app by intensifying interventions influences time on other sites, apps, and devices. We find that on the browser, reducing time on one site reduces time spent elsewhere, but we do not observe the effect on mobile devices, and do not observe cross-device effects. In our third set of studies, we investigate users' motivation levels over time as indicated by the difficulty of interventions they select. We find that users initially overestimate how difficult of interventions they want, and their choices of difficulty progressively decline over time.

Acknowledgments

We thank those who have contributed code, designs, and ideas to HabitLab: Matthieu Rolfo, Sarah Sukardi, Matthew Mistele, Julie Ju Young Kim, Wenqin Chen, Radha Jain, James Carroll, Sara Valderrama, Catherine Xu, Esteban Rey, Lewin Cary, Carmelle Millar, Colin Gaffney, Swathi Iyer, Sarah Tieu, Danna Xue, Britni Olina Chau, Na He Jeon, Armando Banuelos, Kaylie Zhu, Brahm Capoor, Kimberly Te. This work was supported in part by the Stanford Cyber Initiative, as well as by the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program.

This work was supported by a Stanford Human-Centered Artificial Intelligence seed grant. We thank the many users who have used and contributed ideas and feedback to HabitLab.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

We wish to spend our time more productively, but we sink hours into social media; we wish to learn new languages, but we get too busy to practice; we wish to be more healthy, but we do not maintain our exercise routines [31]. Inspired by situations like these, *behavior change* systems help people build new habits and retain them [32, 53, 69, 72]. Behavior change systems draw on theories of persuasion and influence [51, 28] to introduce *interventions*: interaction designs that variously inform, nudge, and encourage people to engage in behaviors more in line with their goals.

There are large numbers of users who wish to achieve behavior change goals, and a large design space for interventions. Thus, there is a natural opportunity to explore the design space of interventions and find what interventions works best, by testing them out with users. However, the existing ecosystem of behavior change tools does not make full use of this resource.

Behavior change tools that are catered towards mass-market adoption by end users tend to employ a one-size-fits-all approach, implementing only a single behavior change intervention and giving the same experience to all users. Users can choose and select between different behavior change apps and extensions to find what they believe works well for them. However, because different apps are developed by different companies which do not share data, they cannot compare the interventions. They thus miss out on a rich opportunity for improving the behavior change systems via experimentation.

Research studies on behavior change, in contrast, have tended to compare only a small number of interventions, with small numbers of paid participants. They thus miss out on the ecological validity, scale, and statistical power that systems targeting the mass market of end users can enjoy.

Our key insight was that we can find an alignment the goals of behavior change researchers and end users, by building a behavior change tool targeted towards the mass market that also runs useful experiments. End users benefit by being able to use a high-quality behavior change tool where the design choices are experimentally tested and validated. Researchers benefit by being able to run ecologically valid behavior change experiments at scale.

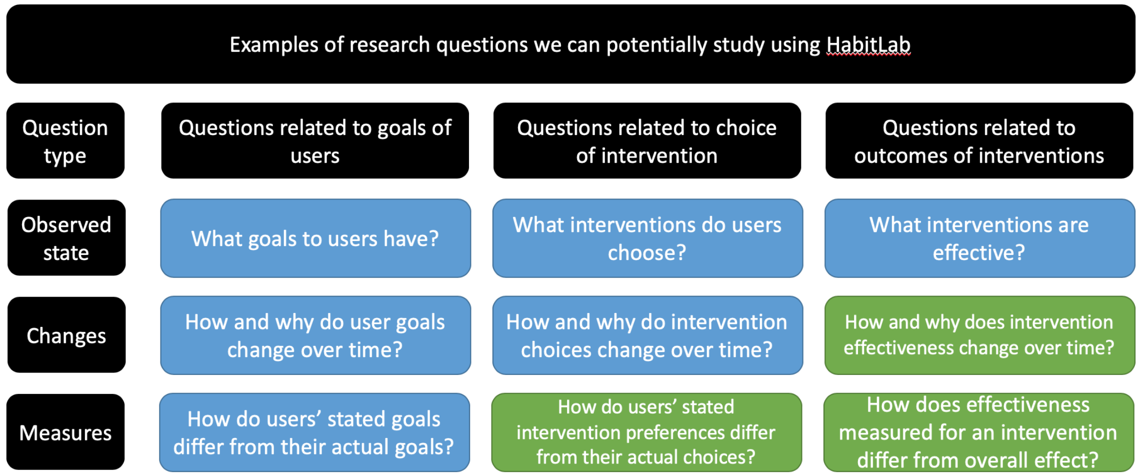


Figure 1.1: Examples of research questions that can be studied with the HabitLab system and the general space of questions they occupy. Research questions we study in this thesis are shown in green.

While we believe that many domains can potentially benefit from in-the-wild behavior change experimentation, we believe that online behavior change is particularly well-suited. As interventions can be distributed as software that requires only a click to install, this enables us to recruit a wide range of participants worldwide for free. Additionally, as our computers and phones can display arbitrary interactive content, this paradigm allows us to experiment with a limitless number of different interventions, and change interventions at any time. Finally, as device usage can be precisely monitored down to the level of which webpage or app was open each second, we can easily measure the effectiveness of interventions and adapt them accordingly.

As a result, we built HabitLab, an in-the-wild experimentation platform for helping users reduce their time online and on their phones. HabitLab is implemented as both a Chrome extension and an Android app, and is currently used by over 12,000 daily active users. Users select sites and apps they wish to spend less time on, and HabitLab deploys a variety of interventions to help them achieve their goals. The platform enables us to run a number of A/B tests comparing interventions and aspects of behavior change systems.

There is a rich set of studies we can run with a tool such as HabitLab. The general paradigm is that users specify goals, interventions are deployed to help them achieve those goals, and we measure the outcomes of the interventions. At a high level, we can thus categorize the space of possible research studies that a platform such as HabitLab can conduct as below. A more detailed classification is shown in Figure 1.1.

1. Studies analyzing users' goals
2. Studies analyzing the choice of interventions to help users achieve those goals
3. Studies analyzing the outcomes of interventions

In this thesis we will present two studies analyzing the outcomes of interventions, and one study analyzing the choice of interventions to help users achieve those goals.

The first study we present asks whether the effectiveness of interventions decline over time. Prior literature suggests this is a possibility, as engagement-boosting novelty effects have been attested in numerous domains. We find that an intervention that is repeatedly presented does indeed decline in effectiveness over time, and that a strategy of rotating between different interventions helps boost effectiveness. This boost comes at the cost of increased attrition, which we find mostly to be due to incorrect mental models, as users are unaccustomed to interventions changing. A simple design helping explain the intervention rotation to users and give them a sense of control significantly reduces this resulting attrition.

The second study concerns itself with whether intervention outcomes are actually what we expect them to be looking at just the time spent on the targeted site or app. Prior literature suggests that willpower is limited, hence we may expect that reducing time on one app, site, or device may increase time spent on others. Other literature suggests that procrastination begets more procrastination by trapping us into a habit loop, hence we may expect that reducing time on one app, site, or device may reduce time spent on others. We find that in the case of site usage on browsers, reducing time on one site results in a reduction of time spent on others. In the case of app usage on mobile, we observe that reducing time on one app does not affect time on others. Likewise, in the case of devices, reducing time on one device does not affect time on others.

The contributions of this thesis are:

1. HabitLab, a system for conducting in-the wild behavior change experiments online. It has been widely adopted with a 12,000 user install base across the browser and mobile platforms, showing that in-the-wild behavior change systems can be used to conduct large-scale experiments.

2. A set of studies conducted on HabitLab showing that static interventions decline in effectiveness over time, and that rotating interventions can improve their effectiveness.

3. A set of studies conducted on HabitLab showing that interventions can sometimes have beneficial secondary effects, reducing time spent on non-targeted sites.

These studies show that our paradigm of in-the-wild experimentation, as realized in the domain of online behavior change via the HabitLab, can work to find novel insights about behavior change systems. We hope this work can help designers build better systems for online behavior change, and promote analogous in-the-wild experimentation in other behavior change domains.

Chapter 2

Related Work

2.1 Behavior Change And Motivation

The field of persuasive technology studies how technology can be used to influence behavior [51]. Persuasive technology systems have been successful in promoting behaviors such as sustainable resource usage [53], fitness [32], sleep [69, 24], healthy eating [97, 42], stress management [3, 117], smoking cessation [99], and productivity [135, 72].

They can operate on many different platforms, such as the web or mobile devices. Web-based systems promote a behavior change goals including classroom engagement [10, 9], psychology therapy [14] and healthy habits [34, 83]. In parallel, a number of studies focused on mobile-based interventions [100, 113, 49, 134, 107]. For instance, MyBehavior, a mobile phone app, was built to track physical activities of the users and to provide personalized suggestions that are tailored to the users' historical behavioral data [107]. Similarly, PopTherapy is a mobile phone app that studied micro-interventions for coping with stress [100].

2.1.1 Theoretical Frameworks for Behavior Change

There are a number of theoretical frameworks describing behavior change systems. B=MAT is a popular framework of behavioral change [51], which demonstrates that systems can focus on three elements—motivation, ability, and a trigger (a call to action)—to produce behavior change. The habit loop is another framework for building habits [45], stating that systems can build habits through an iterated process of displaying a trigger, prompting the user to take an action, giving out a reward, and helping the user to invest in the system.

2.1.2 Taxonomies of Behavior Change

A number of taxonomies characterize the design space of interventions, both general [89, 90, 2, 40] and domain-specific [56, 133]. Michie's behavior change taxonomy lists 93 techniques for behavior change, clustered according to the cognitive phenomenon they target [89]. Systems have investigated effects of these techniques individually, such as using "cheating" to support lapse management [5], using different framings to present results [72], or setting goals and plans [6].

2.1.3 Socialtechnical Systems for Behavior Change

People use a variety of sociotechnical systems to support behavior change, including forums [47, 23], social sharing [103, 27, 102, 73], personal informatics [79, 26], and self-experimentation [68]. People use behavior change forums to gain social support [60] – meeting social needs such as approval and esteem [66]. They do so by providing users with information and advice [60], and establishing norms [23]. They also facilitate social comparisons [37] which influence behaviors, as social comparison theory states that users seek to bring their behaviors in line with norms [48]. Communities also help users find others with similar experiences [61] who can help them through the process of recovering and adapting to changes [94]. Social sharing [103, 111] works by helping users receive support through social interactions, and encouraging accountability [43]. Personal informatics support behavior change through stages of preparation, collection, integration, reflection, and action [79]. The theory of lived informatics [44] adds additional stages where users choose tracking tools, and alternate between lapsing and resuming their tracking behaviors.

2.1.4 Online Behavior Change

One major topic inspiring our work is users' desires to curb or control their time spent on social media sites. People pressure themselves to, and often do, make efforts to reduce their time spent on social media sites such as Facebook and Twitter [121, 118]. Yet this is difficult because users turn to social media to address their need to belong, the need for self-presentation, the need for self-esteem [92], the need for entertainment and gratification [105], and self-affirmation [126]. Whether social media use improves well-being is a complex question depending on the nature of the engagement [130, 84, 81, 71, 91, 115, 124], but thanks to instant gratification and sites' use of gamification [25, 136, 62] and behavior design techniques [51, 45] to drive engagement, users keep coming back to the point that some consider it an addiction [11, 114, 125, 129].

2.1.5 Gamification

Much previous work has focused on gamification as an approach to design behavior change systems [39]. Gamification has been shown to have positive effects on engagement and outcomes in behavior-change contexts such as promoting healthy habits [34, 83] and improving educational engagement [9, 10], though effectiveness varies depending on the context and design [54].

2.1.6 Attrition

Attrition is a major challenge in behavior change systems. Attrition [46], also known as dropout, occurs when participants stop participating, leave, or uninstall the system. Persuasive systems built for weight control and therapy have shown substantial attrition rates in longitudinal studies [19, 100], and prior work in CSCW has sought to help reduce attrition rates through techniques drawn from dieting and addiction research [5].

2.1.7 Personalization

A recent trend in behavior change systems has been the concept of personalizing interventions. Such systems explore several possible strategies using techniques such as multi-armed bandits to find the intervention that is most effective for the user [100, 106]. For example, PopTherapy demonstrated personalized messaging could be found through such techniques [100]. Likewise, HeartSteps conducted tens or hundreds of micro-randomized trials on users [38]. When multi-armed bandits are just beginning to get feedback from a user, they will try out several different interventions to see what works. This exploration has the effect of rotation, but the amount of rotation declines as the bandit begins to personalize. In this paper, we examine the contrarian assertion that perhaps rotation should be maintained to sustain novelty even after the multi-armed bandit is aware of which intervention is most effective for the user.

2.2 Rotating Interventions

In this section, we review literature in behavior change systems and psychology to develop specific testable predictions regarding the research question.

2.2.1 Effectiveness over time

While behavior change systems can be effective [14, 35, 132], many review papers are more restrained in whether behavior change systems remain effective over long periods of time [93, 20, 96, 55]. The critique holds that behavior changes are long, complex processes, and the effectiveness of a system is hard to maintain indefinitely [104]. Prior work suggests that the effectiveness of showing a static intervention cannot be maintained indefinitely [59, 112]. For example, when a health behavior change system started sending email reminders, the first reminder was successful 28% of the time, but by the fifth reminder it was successful only 18% of the time [67]. A further meta-analysis of 88 computer-tailored interventions for health behavior change suggested that the efficacy of interventions decreases over time [76].

2.2.2 The impact of rotation

Novelty can be a driving factor for effectiveness. One study showed that novelty can influence encoding of information into long-term memory, which, in turn, may raise awareness of behavioral changes [74]. Studies of gamification also explore the effect of novelty on user engagement [54].

In web design, people begin ignoring parts of the screen that have little information scent, such as ads. This phenomenon is termed banner blindness, after the commonness of the effect in internet banner advertising [18]. As static interventions remain deployed, they may suffer from the same banner blindness and lack of novelty (wear-out) effects, suggesting a potential mechanism for the decreased effectiveness over time.

Rotating interventions may counter these effects. Different interventions appear in different parts of the interface, making it less likely that the user would ignore them wholesale. Online behavior change systems that use machine learning algorithms such as multi-armed bandits hone in on a small number of interventions to use [100, 106], but during the early exploration phases they are essentially rotating between interventions. Systems that personalize interventions [67] or deploy many micro-studies [38] have generally found positive effects.

2.2.3 Attrition

Attrition is a major challenge in behavior change systems: a metastudy of eHealth interventions found that an attrition rate around 99% over a 12-week period is normal [46]. Likewise, the number of users in a stress-coping mobile application declined in a steady rate through the study [100].

Though rotating interventions aids novelty, the literature suggests that it may hurt attrition. Rotation violates usability heuristics such as consistency and user control [95]. Specifically, users may perceive a loss of control when they are presented with ever-changing interventions, potentially leading to non-compliance behaviors and a higher attrition rate [58]. Typically, in attrition-risky domains such as education, an effective user-centered design is critical for minimizing attrition [12].

2.3 Measuring Overall Effectiveness

Measuring the effectiveness of a persuasive system remains a major challenge in the design of behavior change systems. While behavior change systems can be effective during experiments [14, 35, 132], many review papers are more restrained in whether behavior change systems remain effective outside studies and bring longitudinal behavioral change [93, 20, 96, 55]. Because behavior changes are long and complex processes, the efficacy of a persuasive system is often difficult to measure [104]. For instance, an intervention promoting healthy habits, which was effective in changing participants' eating habits, might reduce their physical activities, which were not measured in the experiment [33]. Likewise, a system promoting increased physical activity may be unable to observe effects on participants' eating habits [38]. Compared to prior work, our study examines these spillover effects in the context of a more complete ecosystem, including both desktop browsers and mobile devices.

2.3.1 Multi-Device Usage

Cyberslacking, referred to as non-work-related computing, is the use of Internet and mobile technology during work hours for personal purposes [131, 101, 65, 80]. One study found that employees spent at least one hour on non-work-related activities during a regular work day [131]. Researchers also reported that non-work-related Internet usage comprises approximately 30%–50% of total usage [1, 63].

Unproductive time begets further unproductive time. For example, increased time spent online can increase sleep debt, which in turn leads to more time spent online [88]. Likewise, the Hook Model claims that many of the most addictive online sites use a cycle of investment techniques to keep users coming back—for example, making a post on Facebook may result in future notifications, which will in turn will get the user to come back and make more posts [45]. Finally, sites such as Facebook, Reddit, Twitter, and Buzzfeed are filled with links to each others' content, so it may be the case that increasing usage of one will increase usage of others. If productivity interventions are able to break this vicious cycle of procrastination for one application, they may actually reduce time spent on other unproductive applications as well.

2.3.2 Distribution Of Unproductive Time

In this section, we will examine related studies in behavior change systems to develop testable hypotheses regarding the research question.

Multitasking has become ubiquitous in today's workplaces [13, 85, 22]. Multitasking is both essential and unavoidable in the workplace [52, 86], and it takes 11 minutes on average before people switch to a new task [36].

Studying behavior change effects across multiple devices is important: focusing on a single platform will myopically miss unproductive behaviors on other platforms. Attention is fragmented in both mobile and traditional desktop environments [78, 85]. The time spent on mobile devices has increased more rapidly than time on computers or TVs [17, 30]. On the other hand, mobile applications have been regarded as substitutions of websites in many studies [127]. Large technology companies such as Facebook and Amazon have been focusing on user growth on mobile devices [78].

However, interventions may result in unintended outcomes [57, 55, 123]. Specifically, while some interventions may be highly effective at achieving the measured goal of a behavioral change system, they may reduce desired outcomes elsewhere [55]. In one health-related intervention, while the physical activity of participants increased, calorie intake also increased, working against the goal of promoting a healthy lifestyle [21]. Similarly, using peer pressure to build confidence for students at school would, in turn, lower their self-esteem which actually was opposite to the goal of augmenting confidence [123].

Chapter 3

The HabitLab Behavior Change Experimentation Platform

3.1 Introduction

Studying behavior change requires in-the-wild intervention and observation [32]. Inspired by previous CSCW tools for naturalistic data collection [109], we developed HabitLab [75], an open-source[1] platform, as a living laboratory to help us understand online behavior change and as a platform to explore novel behavior change designs (Figure 3.4).

HabitLab is a Chrome browser extension and Android application that contains a variety of productivity interventions. It aims to help users reduce their time spent online on web pages that the user specifies (e.g., Facebook, Twitter, and Reddit). The system is pitched to end users as a tool that explores various different interventions (referred to as "nudges") to help them reduce their time on sites.

Both versions follow the structure of allowing users to choose what they wish to spend less time on (setting goals), and deploying interventions to meet those goals. On the Chrome version, users choose sites to spend less time on (goal sites – for example, `facebook.com`), as shown in Figure 3.2. On Android, users choose particular apps to spend less time on (goal apps – for example, the Facebook Android app), as shown in Figure 3.1. Interventions are deployed when users visit a goal site on Chrome (Figure 3.3), and when users open a goal app on Android, as shown in Figure 3.1.

Users install the extension, and go through an onboarding process where they select sites and apps they wish to reduce their time on (Figure 3.5). There are predefined options—Facebook and YouTube are selected by default, as they were the most commonly used—but users can also add any custom site. Custom sites are suggested via an analysis of the user's browsing history. The system explains to users that they will be shown a variety of interventions (Figure 3.6), a form of self-experimentation [68], to help them reduce

[1]HabitLab is available at `http://habitlab.github.io`.

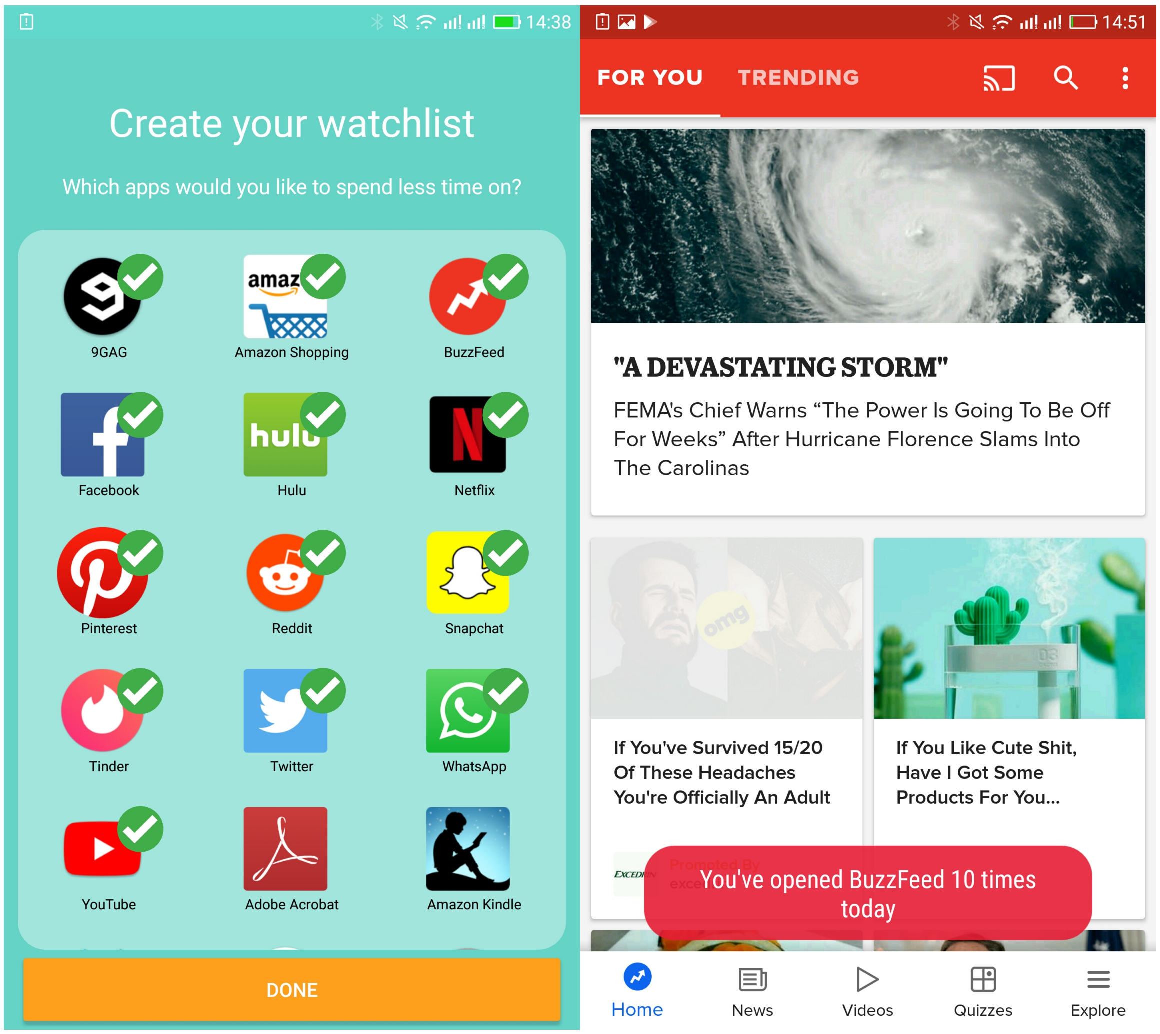


Figure 3.1: Screenshots from the mobile version of HabitLab.
Left: The goal selection screen, where users choose which apps to spend less time on.
Right: An example intervention, which shows the visit count when a user opens a goal app.

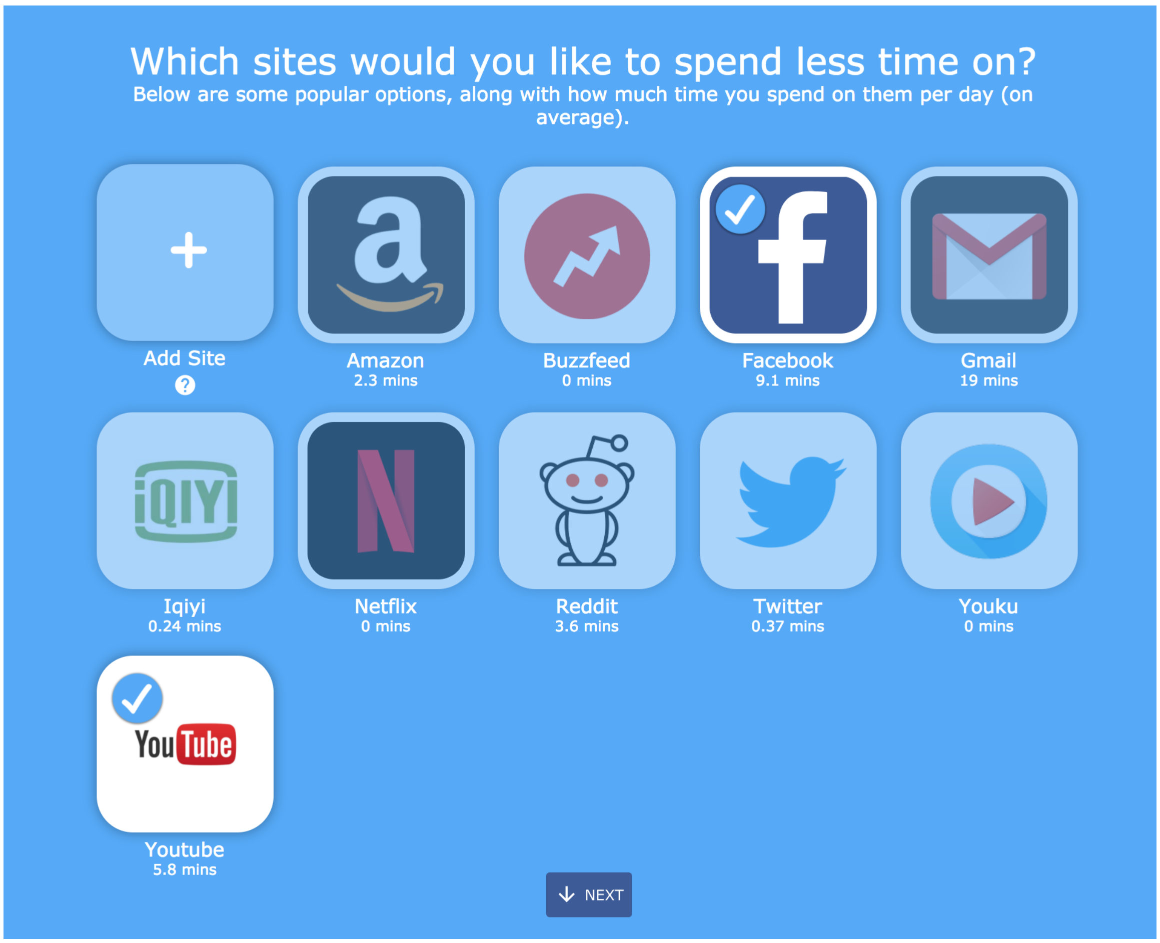


Figure 3.2: The goal selection screen, where users choose which sites to spend less time on (browser version).

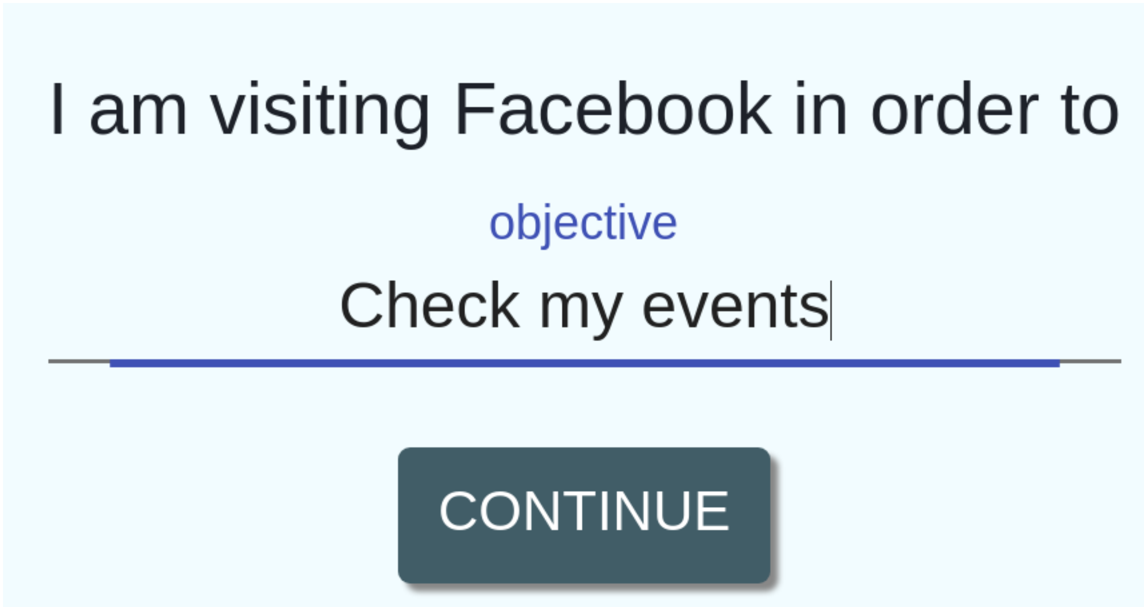


Figure 3.3: An example intervention, which asks a user to write their objective for visiting a site (browser version).

Figure 3.4: HabitLab's homepage describes the browser extension and mobile application. Users adopt it to try out a large number of different possible interventions, called nudges.

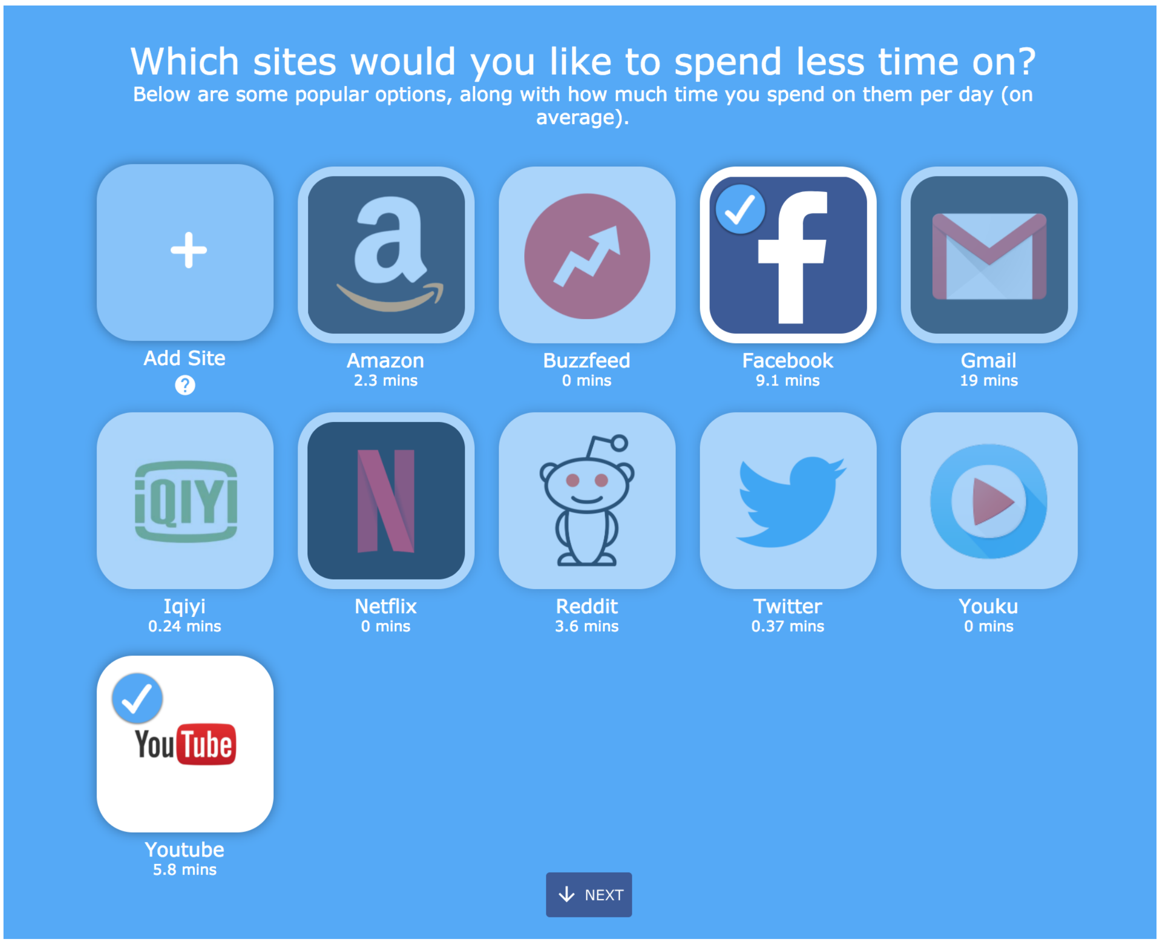


Figure 3.5: During onboarding, users choose which sites they want to spend less time on.

Figure 3.6: Users are presented with the interventions they will see on each site.

time on that site. These interventions are typically targeted to each site, for example a news feed blocker for Facebook or a related video hider for YouTube. However, some interventions such as a stopwatch timer can be added to any custom site. Users can preview the interventions for the sites they select, and enable or disable each intervention if desired. Users can later enable or disable interventions and sites through a settings page.

HabitLab emphasizes to users the availability of multiple interventions and that it may show users different interventions each time they load a page. This emphasis is made clear on the HabitLab website, Chrome store listing, and through features in the dashboard such as visualizing the relative effectiveness of different interventions. HabitLab implements a multi-armed bandit algorithm to explore and find the interventions that are most effective for each user, optimizing for minimizing time spent on a site. However, in the experiments in this paper, we disabled this functionality and instead used simple random selection so that we can study the effects of rotation in isolation.

3.2 Mobile and Browser Versions

The Chrome extension and Android app differ in some minor details. They support different sets of goals: users select apps to reduce time on in the Android version, whereas users choose sites to reduce time on in the Chrome version. Additionally, the specific set of interventions available differs between the platforms to fit the design languages of the browser and the mobile phone. The Chrome version has certain interventions which are site-specific – such as a news feed remover that is specific to Facebook. However, because Android does not allow applications to edit each other's view trees, the Android version's interventions are all glass pane overlays, and thus are general and can be used on any app. The concept of a session is different on the platforms: in the Chrome version, a session is time on a site until that tab is either closed or the user goes to a different domain. Time measured is active time – so if the tab is not focused, or if there is no keyboard or mouse activity for over a minute, the timer is temporarily paused. However, on Android, because there is no concept of a tab, the measurement of a session is different. There, a session is considered the duration over which an app is opened and focused. Closing the app, switching to a different app, or turning off the phone will end the current session.

3.3 Design of HabitLab Interventions

HabitLab can track time and deploy interventions on all sites, but some interventions are tailored towards specific sites. There are 27 interventions total: seven generic interventions that can be used on all sites, five interventions designed specifically for Facebook, and additional ones designed specifically for YouTube, Reddit, Twitter, Netflix, Gmail, Amazon, iQiyi, and Youku.

Interventions are designed drawing on theories of behavior change—for example, goal setting theory [82], persuasion [28, 51, 2], and gamification [39]. A sample of the interventions available for Facebook, categorized according to underlying strategies and theories, are shown in Table 3.1. Screenshots of some Facebook interventions are shown in Figure 3.7. Descriptions of the interventions on the Chrome and Android versions can be found at the end of this chapter.

Not all interventions are enabled by default—this is because some of them have higher attrition rates than others. Non-default interventions can be previewed and enabled by users during onboarding and on the settings page. The interventions enabled by default were the ones we found to have low attrition rates during pilot deployments—we chose this strategy to ensure user retention and growth, which is a prerequisite for gathering data in an in-the-wild experiment setting.

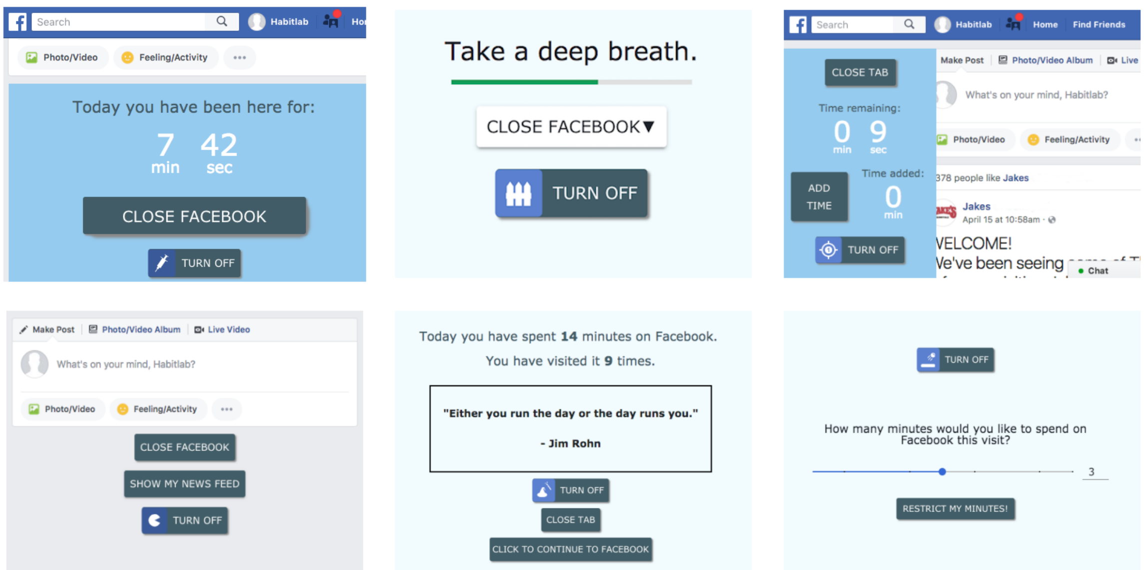


Figure 3.7: Examples of interventions available for reducing time on Facebook. From left to right, top to bottom: a timer injected into the news feed; a page before opening Facebook requiring that the user wait a few seconds before visiting; a countdown timer that automatically closes the tab after time elapses; an opt-in required to show the news feed; an interstitial page before opening Facebook with a quote; an interstitial page before opening Facebook that requires the user set a time limit for how long they will spend this session.

Strategy	Theory	Intervention
Commitment	Self-consistency theory [8, 28, 120]	Ask the user to set a goal for the length of time they will stay on the site (generic)
Enforce default limits	Status quo bias [116]	Automatically close tab after 60 seconds unless the user clicks a button to ask for more time (generic)
Reduce social incentives	Social proof [119, 28]	Hide Facebook comments by default (default)
Delaying Rewards	Operant conditioning [15]	Make the user wait 10 seconds before visiting Facebook (generic)
Removing Rewards	Operant conditioning [15]	Hide the news feed (default)
Inform the user	Theory of reasoned action [7]	Show a counter at the top of the page of how long user has been on Facebook today (default, generic)

Table 3.1: A subset of the interventions for Facebook, categorized according to persuasion strategy and theory. Interventions that are enabled by default are marked *default*, interventions that are available for all sites are marked *generic*.

3.4 HabitLab adoption and user demographics

As of writing, the browser version of HabitLab has over 12,000 daily active users, and the Android version has over 500 daily active users.

Demographics according to Google Analytics indicate that our users are 81% male, with the most commonly represented age group being 25-34, as shown in 3.8.

Our userbase represents a diverse set of countries and languages – users represent 151 countries as shown in 3.9. The top 10 countries are shown in 3.10 – the US is the most-represented country, representing 30% of the userbase.

Half of our userbase uses English as their preferred language for displaying webpages, as indicated in 3.11. Volunteers have translated HabitLab into 13 languages (Arabic, Chinese, Czech, Dutch, French, German, Greek, Italian, Polish, Portuguese, Russian, Spanish, and Turkish).

The users were not explicitly recruited, but were rather all organic installs who discovered the extension/app via sources such as the Chrome/Play store, or were referred to it via press coverage in sources such as Wired or the New York Times.

Users are asked to read and provide consent to the research protocol upon installation. They may opt out of data collection if they do not wish to have their data analyzed for research purposes.

3.5 Design principles and tradeoffs

We designed HabitLab from the start intending it to be a in-the-wild experimentation platform with a large number of users who would voluntarily and organically install it. As a result, we made a number of design decisions that prioritize growth and retention.

Interventions can all be disabled by the end user, either temporarily for the duration of a session via a

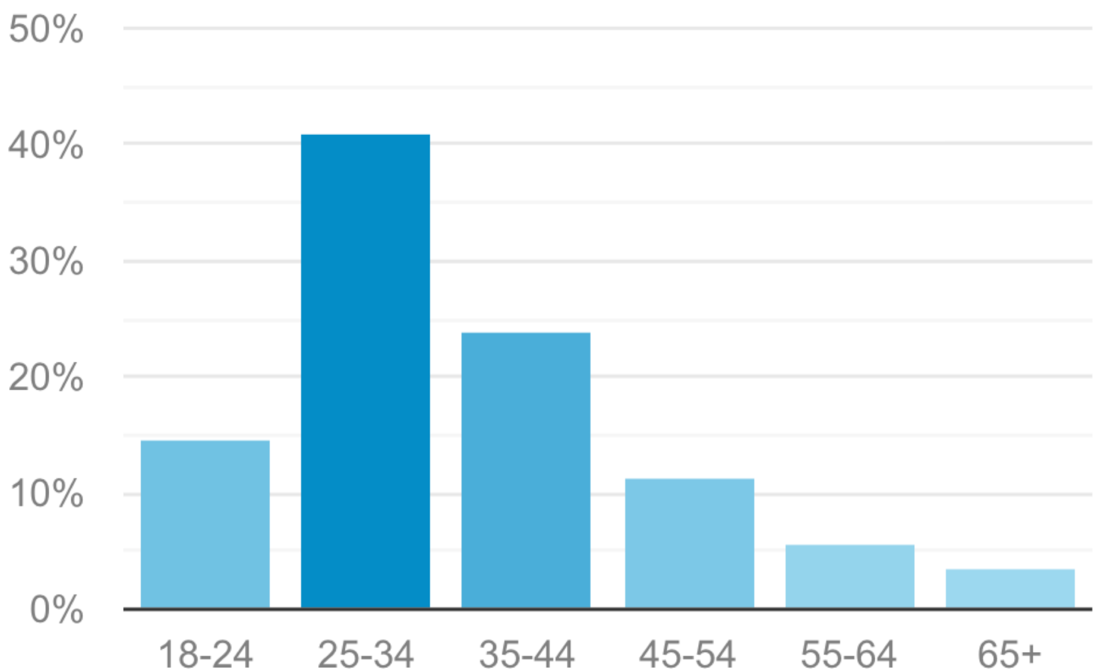


Figure 3.8: Ages of HabitLab users. 25-35 is the most-represented demographic.

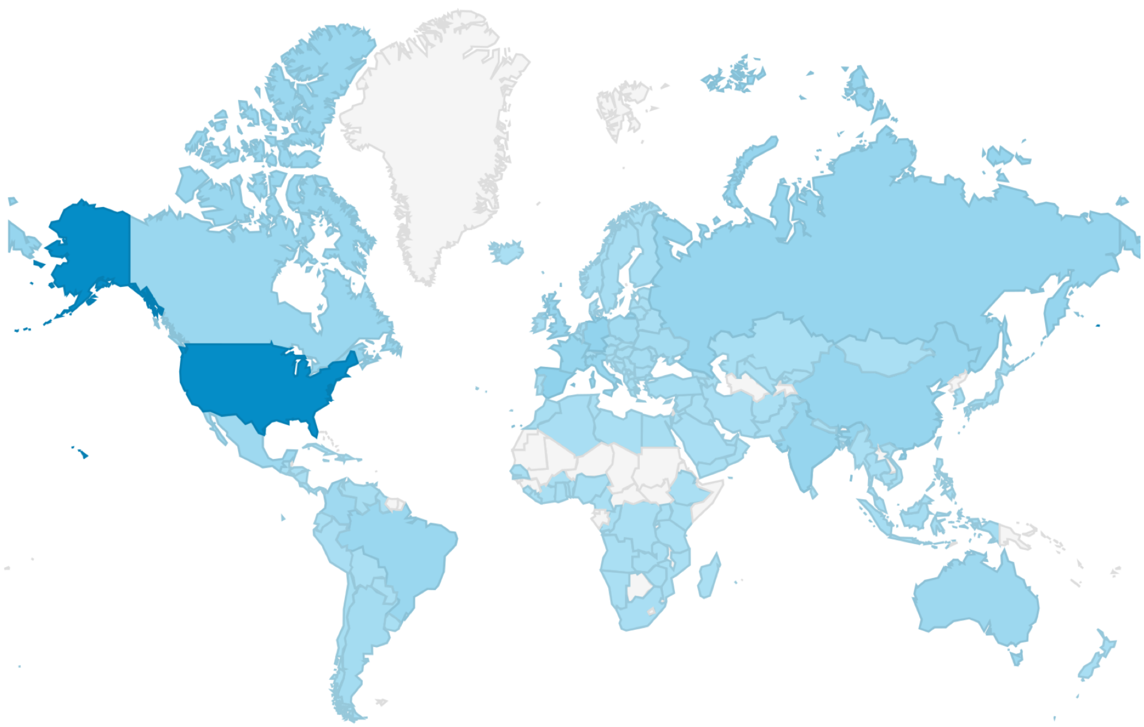


Figure 3.9: Map of countries representing HabitLab users. North America, Europe, and Asia are all well-represented.

1.	United States	3,710	30.16%
2.	Spain	605	4.92%
3.	Germany	540	4.39%
4.	Russia	461	3.75%
5.	China	450	3.66%
6.	India	448	3.64%
7.	United Kingdom	439	3.57%
8.	France	418	3.40%
9.	Italy	402	3.27%
10.	Canada	368	2.99%

Figure 3.10: Top 10 countries using HabitLab. Users from the US account for 30% of our userbase.

1.	English (US)	6,143	50.00%
2.	Spanish	801	6.52%
3.	English (UK)	725	5.90%
4.	Chinese (Simplified)	462	3.76%
5.	Russian	443	3.61%
6.	Italian	340	2.77%
7.	Portuguese (Brazil)	293	2.38%
8.	French	289	2.35%
9.	German (Germany)	260	2.12%
10.	German	212	1.73%

Figure 3.11: Languages that HabitLab users set as their preferred language to show webpages in. English is the preferred language of half of our users.

"Turn off" button shown on each intervention, or permanently. This is intended to boost retention by preventing uninstalls caused by users disliking a particular intervention. While this complicates some analyses – for instance, we may have fewer samples about the effectiveness of less popular interventions that tend to be disabled more – we believe this to be the appropriate tradeoff.

Interventions are designed to be minimally intrusive. While in principle users can just disable interventions they do not like, we still found that many users would uninstall after seeing particular, intrusive interventions. We saw this pattern most notably with interventions that have interstitial screens – that is, they prevent the user from interacting with the page until they have gone through the intervention. As a result, with the exception of a handful of interventions which must be in the interstitial format – for instance, forcing the user to wait for 10 seconds before loading the page – we tried to avoid interstitial interventions as much as possible.

Interventions are designed to load fast, and as a result we ensure that all interventions can work offline and do not depend on remote network resources that might take a long time to load. We found that for interventions that take a second to load or more, the uninstall rate would increase after seeing them. This effect is particularly evident with interstitial interventions, which would have the jarring effect of allowing the user to use the site for a few seconds and disrupting them with an interstitial page once the intervention loads.

We have a simple and short onboarding process. Notably, we do not have long demographic surveys that characterize other similar research projects like LabInTheWild, relying on data from Google Analytics to gather demographic data instead. While data from Google Analytics is only approximate – it is estimated from browsing patterns rather than from asking users directly – we believed that if the data is accurate enough to be used by market research companies worldwide, it would be adequate for our purposes. Furthermore, requiring users to complete onboarding demographic surveys would not be able to guarantee that users would answer truthfully.

This principle of minimizing the amount of questions the user must answer also extends beyond the onboarding process. We make only minimal usage of experience sampling. We do so because we saw in one of our studies that even minimally intrusive, single-click experience sampling prompts that users can safely ignore will significantly increase the uninstall rate. As a result, most of the data we are able to gather is quantitative in nature, and we are only able to gather limited qualitative data from what users report to us through email, our feedback pages, reviews left on app store pages, or the uninstall survey.

3.6 List of Browser Interventions

The following is the list of interventions used for this study, showing the intervention name and description as seen by the end user.

Generic interventions that can be used on all sites:

- Minute Watch: Notifies you of time spent every minute

- Supervisor: Shows time spent on site at the top of screen
- Scroll Freezer: Freezes scrolling after a certain amount of scrolls
- Stat Whiz: Show time spent and visit count each visit
- GateKeeper: Makes you wait a few seconds before visiting
- 1Min Assassin: Closes tab after 60 seconds
- Bouncer: Asks how long you want to spend on site this visit

Facebook-specific interventions:

- Time Injector: Injects timer into the Facebook feed
- Feed Eater: Removes the Facebook news feed
- TimeKeeper: Notifies you of time spent in the corner of your desktop
- No Comment: Removes Facebook comments
- Clickbait Mosaic: Removes clickbait from the news feed

Youtube-specific interventions:

- Sidekicker: Remove sidebar links
- Think Twice: Prompt the user before watching a video
- No Comment: Removes comment section

Netflix-specific interventions:

- Fun Facts: Gives you a fact and links an article on the effect of TV
- Alarm Clock: Asks the user to set an alarm before watching a show
- Stop Autoplay: Stops the site from automatically playing the next video

Reddit-specific interventions:

- Comment Remover: Removes Reddit comments
- Mission Objective: Asks what you aim to do this visit and puts a reminder up

Youku-specific interventions

- Think Twice: Prompt the user before watching a video
- Sidekicker: Remove sidebar links

iQiyi-specific interventions

- Think Twice: Prompt the user before watching a video
- Sidekicker: Remove sidebar links

Twitter-specific interventions:

- Feed Eater: Removes the Twitter news feed

Amazon-specific interventions:

- No Recs: Hides recommendations

Gmail-specific interventions

- Speedbump: Delays the arrival of new emails

3.7 List of Mobile Interventions

All mobile interventions are generic, that is they can be used on any app.

- At it Again: Sends a pop up with your app visit count.
- Progress Report: Sends a pop up with today's total usage for a certain app
- Red Alert!: Sends a notification with today's total usage for a certain app
- Repeat Offender: Sends a notification with your app visit count
- All in All: Pops a dialog with the day's total time on the current app
- Back To Target: Suggests you to visit a target app
- Counting on You: Puts a timer on screen in watchlisted apps
- Man Overboard! Shows a dialog with your app visit count
- No Peeking!: Asks for confirmation before opening watchlisted apps
- Wait Up! Pause for 10 seconds before entering an app
- Your Better Half: Sends a pop up to go to a target app
- Look on the Bright Side: Dim the screen a little at a time
- Take Your Pick: Select how long you want to spend on an app
- The Final Countdown: On screen timer that closes the app when time runs out

The following interventions apply across the device as a whole, not individual applications.

- How Time Flies!: Sends a pop up message with current app visit length
- Knock Knock: Sends a pop up with your glance count for the day
- Long Time No See: Sends pop up with your phone usage for the day
- Call it a Day: Sends notification with phone usage for the day
- Easy on the Eyes: Sends notification with glance count for the day
- Hello, Old Friend: Sends notification with unlock count for the day
- The Clock is Ticking: Sends a notification with the current app visit duration
- En Garde: Pops a dialog with the day's total unlock count
- Hold the Phone: Show dialog with phone usage for the day
- Long Story Short: Pops a dialog with the visit time for the current app
- Quote reminder: Show quote upon opening app
- Time Reminder: Show dialog with phone usage for the day
- Take Your Pick: Select how long you want to spend on an app

Chapter 4

Rotating Online Behavior Change Interventions

4.1 Introduction

Behavior change systems today suffer from declining effectiveness as novelty wears off over time. Typically, behavior change systems utilize a *static* design, which never changes. For example, to manage social media browsing time, the three popular options are to block tempting sites [128], use a work timer [50], and audit time spent [70, 110]. However, static interventions suffer from high attrition and abandonment rates [29, 46], and interventions decline in effectiveness over time [76]. Habituation eventually drives users to stop paying attention to, or avoid, static interventions—an effect often seen on the web as banner blindness [18]. The end result is that many behavior change systems are tuned out by users, and are unsuccessful at their goals.

If static interventions are tuned out, *rotation* might provide a remedy. Much like a human coach or tutor rotates between different approaches over time, rotation might maintain attention in ways that static interventions cannot. Online behavior change tools could apply similar techniques, for example injecting a visible stopwatch timer into the page on one visit to Facebook, and hiding comments on the next visit. Techniques that personalize interventions via multi-armed bandits show positive treatment effects, suggesting that the approach may hold promise, but this existing work cannot separate the effects of personalization from the effects of rotation [108, 77, 100]. Because rotated interventions continually change the user interface, however, they may frustrate users by violating consistency and a sense of user control, leading to lower effectiveness or higher attrition.

This paper takes up the question: are static or rotated interventions more effective for behavior change? Is it possible to understand the effects of rotation in order to design more effective behavior change systems? We focus specifically on helping users who want to manage their time on social media websites such as Facebook, YouTube, Reddit, and Twitter. We perform a series of field experiments with people who sought out and installed a browser extension that we developed for online behavior change.

Our platform, *HabitLab* (`https://habitlab.stanford.edu`) is a Chrome extension that features a number of online productivity interventions to help users reduce their time spent on sites such as Facebook. We released HabitLab publicly on the Chrome web store, where it has attracted over 8,000 daily active users. This user base allows us to observe real-world usage and attrition patterns over time.

We ran three in-the-wild studies on users who newly installed HabitLab. In Study 1, we compared static interventions to rotation. We measured effectiveness through time on the user's targeted site, and we measured attrition by tracking when users stopped using the extension. Results indicate that rotation is a double-edged sword. Rotating interventions reduced time spent on sites by 34% per day, but at the cost of nearly doubling attrition levels.

Study 2 replicates the first experiment over a longer period of seventy days, and additionally tests whether the number of interventions included in the rotation impacts attrition. The results successfully replicated the original results over a longer 70-day period, and suggested that the larger the set of interventions, the higher the probability of attrition.

To investigate the underlying causes of attrition and mitigate the effects of rotation on attrition, we analyzed user feedback and developed a pair of interface techniques to improve the user experience in the presence of rotation, which we deployed in Study 3. The first technique is informational, aiding people's mental models by reminding them that the system may show a different intervention on each visit. The second technique focuses on user control, providing the same information as well as a just-in-time mechanism for people to opt out of each new intervention as they see it. Results indicated that these interventions reduced attrition by over half, so that 80% of new users were still using the system actively after a week.

In sum, this paper contributes the first comparison of static and rotation intervention strategies in behavior change, a living laboratory system that allows us to deploy this investigation and other field experiments, and interaction design strategies that can help offset increased attrition due to rotation. Its results suggest that people may be more able to control their social media usage more effectively than using today's common techniques such as site blockers. The rest of the paper is organized as follows: we first review studies of behavior change to develop our research question and hypotheses; we then describe our studies and results; we close with reflections and future design directions.

xcessive recreational online activity, or "cyberslacking" is a major problem. Hence, numerous productivity interventions, such as time trackers and site blockers, have emerged to help users reduce their time online. Existing productivity tools typically always show the same intervention; however, techniques that alternate between different interventions (selected via algorithms such as multi-armed bandits) have been found to improve overall effectiveness in various health intervention contexts. This can be applied to online productivity tools as well – i.e., hiding the news feed on one visit, injecting a timer into the page on others, etc.

Alternating interventions has typically been done in the context of an adaptive intervention selection algorithm (e.g. multi-armed bandits), and the benefits have been attributed to personalization – certain interventions are more effective for certain individuals, and by exploring which interventions are most effective for the individual, the algorithm can maximize effectiveness in the long run.

However, we believe that our understanding of the effects of alternating interventions may be incomplete. Certain health interventions have been found to decline in effectiveness over time, and banner blindness is a well-known phenomenon in online advertising – if users analogously develop blindness to interventions, could the novelty brought by alternating interventions itself improve effectiveness, by preventing the user from developing intervention blindness? Such effects are important, because declines in intervention effectiveness over time violates a major assumption of multi-armed bandit algorithms – that the rewards, e.g. the effectiveness of the interventions – do not change over time. If alternating interventions has a significant positive effect on effectiveness, intervention selection algorithms need to take this into account.

Alternatively, alternating interventions can have negative effects – users may be annoyed and fatigued if they constantly see different interventions, as it may violate expectations of a consistent user experience. Users may also feel a lack of control if the system is choosing which interventions to show them, rather than always showing the same one. If a user particularly dislikes one of the interventions, the user may be exposed to it while the system is exploring the space of interventions, which risks causing user dissatisfaction. These factors may lead to alternating interventions contributing to increased rates of attrition – which intervention selection algorithms need to take into account.

To answer these questions of in which ways alternating interventions can be beneficial or harmful, we developed a Chrome extension, HabitLab, that features a number of online productivity interventions to help users reduce their time spent on sites like Facebook. We released it publicly on the Chrome web store, and ran a pair of in-the-wild studies on users who newly installed the extension, where we varied how much we were alternating between interventions. Users were unpaid, organic installs from the Chrome web store rather than artificially recruited participants, which allows us to better observe real-world usage and attrition patterns. We found the following results:

1. Users who are constantly rotating between interventions have higher rates of attrition than users who consistently see the same intervention.

2. Interventions decline in effectiveness over time

3. Interventions are more effective at reducing users' time spent when they are shown alternating between different interventions

4. Users are less likely to attrition if they chose their interventions themselves during the onboarding process

These results led us to hypothesize that the increase in attrition observed due to alternating interventions may be a result of violation of user expectations (due to users having an incorrect mental model), or users feeling a lack of control.

To investigate the underlying cause and mitigate this attrition, we developed a pair of interface techniques shown when a new intervention is seen for the first time. The first interface just repeats information (that users saw during onboarding but may not have read) that the system may show a different intervention on each visit, while the second interface gives users a sense of control by allowing them to opt-out of seeing the intervention in the future. We found that adding these interventions both significantly decrease attrition (fake result).

4.2 Research Questions

A recent trend in behavior change systems has been the concept of personalizing interventions. Such systems explore several possible strategies using techniques such as multi-armed bandits to find the intervention that is most effective for the user [100, 106]. For example, PopTherapy demonstrated personalized messaging could be found through such techniques [100]. Likewise, HeartSteps conducted tens or hundreds of micro-randomized trials on users [38]. When multi-armed bandits are just beginning to get feedback from a user, they will try out several different interventions to see what works. This exploration has the effect of rotation, but the amount of rotation declines as the bandit begins to personalize. In this paper, we examine the contrarian assertion that perhaps rotation should be maintained to sustain novelty even after the multi-armed bandit is aware of which intervention is most effective for the user.

The challenges of static interventions, and the rising wave of personalization systems, call into focus: would a rotation strategy work? Or is it a weak palliative with little discernible effect? This led to our research question:

Research Question (RQ)**.** *Can a strategy of rotating interventions produce more effective behavior change systems?*

While behavior change systems can be effective [14, 35, 132], many review papers are more restrained in whether behavior change systems remain effective over long periods of time [93, 20, 96, 55]. The critique holds that behavior changes are long, complex processes, and the effectiveness of a system is hard to maintain indefinitely [104]. Prior work suggests that the effectiveness of showing a static intervention cannot be maintained indefinitely [59, 112]. For example, when a health behavior change system started sending email reminders, the first reminder was successful 28% of the time, but by the fifth reminder it was successful only 18% of the time [67].

A further meta-analysis of 88 computer-tailored interventions for health behavior change suggested that the efficacy of interventions decreases over time [76]. This prompts our first hypothesis:

Hypothesis 1 (H1)**.** *Static interventions will suffer from decreased effectiveness over time.*

Novelty can be a driving factor for effectiveness. One study showed that novelty can influence encoding of information into long-term memory, which, in turn, may raise awareness of behavioral changes [74]. Studies of gamification also explore the effect of novelty on user engagement [54].

In web design, people begin ignoring parts of the screen that have little information scent, such as ads. This phenomenon is termed banner blindness, after the commonness of the effect in internet banner advertising [18]. As static interventions remain deployed, they may suffer from the same banner blindness and lack of novelty (wear-out) effects, suggesting a potential mechanism for the decreased effectiveness over time.

Rotating interventions may counter these effects. Different interventions appear in different parts of the interface, making it less likely that the user would ignore them wholesale. Online behavior change systems that use machine learning algorithms such as multi-armed bandits hone in on a small number of interventions to use [100, 106], but during the early exploration phases they are essentially rotating between interventions. Systems that personalize interventions [67] or deploy many micro-studies [38] have generally found positive effects.

Based on these results, non-static interventions may be effective. We hypothesize:

Hypothesis 2 (H2)**.** *Rotation will increase effectiveness, compared to static interventions.*

Attrition is a major challenge in behavior change systems: a metastudy of eHealth interventions found that an attrition rate around 99% over a 12-week period is normal [46]. Likewise, the number of users in a stress-coping mobile application declined in a steady rate through the study [100].

Though rotating interventions aids novelty, the literature suggests that it may hurt attrition. Rotation violates usability heuristics such as consistency and user control [95]. Specifically, users may perceive a loss of control when they are presented with ever-changing interventions, potentially leading to non-compliance behaviors and a higher attrition rate [58]. Typically, in attrition-risky domains such as education, an effective user-centered design is critical for minimizing attrition [12]. In light of these results, we hypothesize:

Hypothesis 3 (H3)**.** *Rotation will increase attrition, compared to static interventions.*

4.3 Experiment Platform: HabitLab

We conducted the studies in this chapter using the browser version of HabitLab. At the time the studies presented in this chapter were conducted, there were 8000 daily active users using the platform. The list of interventions that were included in HabitLab at the time of this study is included at the end of this chapter.

4.4 Study 1: Field Study on the Effect of Rotating Interventions

Our first study is a within-subjects design run on the HabitLab platform that aims to understand the effects of rotating interventions on effectiveness and attrition.

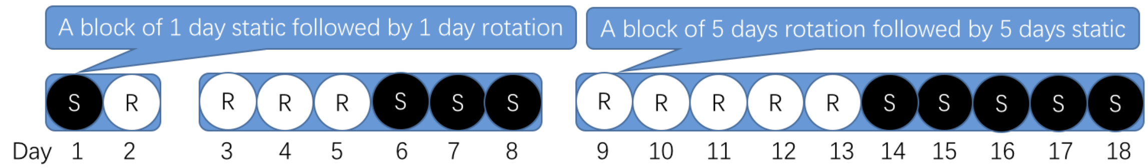


Figure 4.1: Example order in which a user might see conditions. Each circle represents a day – on black days the user is in the "static" condition, white is the "rotation" condition. The order of blocks is randomized; here, this participant is seeing blocks in order 1, then 3, then 5, then 7 (omitted in the figure).

4.4.1 Participants

Participants in our first study consisted of new HabitLab users installing the system over a period of three weeks in March and April 2018. 692 users installed HabitLab over the course of our experiment and consented to our research protocol. We discarded participants who were not new users of HabitLab, since some users were reinstalls or new devices for existing users. We also discarded participants who did not the complete the onboarding process, or who uninstalled the system before they saw their first intervention. This left us with 217 participants.

We do not administer a demographic survey at install time, because long onboarding processes had previously led to high abandonment. Many users find HabitLab through routes other than the web site, but Google Analytics on the web site can provide some window into rough trends. Google Analytics estimates that 89% of visitors to the HabitLab website during the experiment period were male, indicating a male skew. The most common age group was 25–34 (41%), followed by 18–24 (29%), 35-44 (22%), and 45–54 (7%). According to users' IP addresses, the most highly represented countries were the US (23%), India (12%), Germany (9%), France (5%), and the UK (4%).

Participants agreed to our informed consent protocol during onboarding. This consent protocol indicated that HabitLab would be selecting and rotating between different interventions, but did not mention any specific algorithm or rotation schedule.

4.4.2 Method

Participants used HabitLab in the course of their usual web activity. As they browsed, HabitLab would introduce interventions when appropriate. All interventions were available to all conditions, but the pace at which old interventions were replaced by new ones depended on condition. Users would react to the intervention, or not, as they browsed.

HabitLab operated on all web sites that the user had selected upon installation. However, because users spend differing amounts of time on different domains, and there was a long tail of domains which were set as goal sites by only a few users, we restricted analysis to domains where we had a substantial dataset, specifically: Facebook, Youtube, Reddit, Twitter, VK, and Amazon.

The experimental unit was a participant assigned to a condition for a block of days. Block lengths were randomized between 1 day, 3 days, 5 days, and 7 days, in order to give us insight into the effects of rotation strategy over different time horizons. When participants were randomized to a block, for example a five-day block, they experienced HabitLab in one condition for five days, then in the other condition for five days, for a total of ten days (Figure 4.1). Condition order was randomized within each block. At the conclusion of a block, the user was then moved into another block length and the trial repeated. The sequence of block lengths was randomized for each participant. If they kept the system installed, participants would experience all blocks after 32 days.

4.4.3 Conditions

We developed a within-subjects repeated measures design, where users alternated between blocks of time during which they were shown either a static intervention or rotated interventions. A within-subjects design such as this allows us to better control for the large variability across users in how much time they spend on a site.

The static condition captures a typical behavior change design with one strategy. At install time, for each site, the participant is randomly assigned a single intervention among the ones that are enabled for the site. Whenever they visit that site on a day in the static condition, the participant will always see that intervention, i.e. the static intervention is the same across all blocks.

The rotation condition captures a strategy of keeping the interventions changing so that users do not begin ignoring them. Each time a participant in the rotation condition visits a target site (e.g., Facebook), HabitLab picks a random intervention from the enabled set to display.

So, in a five day block, a user might spend five days in the static condition seeing the same intervention each time, then five days in the rotation condition seeing randomly selected interventions each time. They are then moved into another block and the method repeats.

4.4.4 Measures

We measured the *effectiveness* of the system as the number of seconds spent on the target site each day. Time, of course, does not perfectly correspond to attention or engagement behavior, as users can get distracted and not actually attend to a web page. However, prior work has generally found it to be an effective estimate (e.g., [135]). To determine whether the user is actively using a target site, we use Chrome's internal definition of active — the browser window and tab is focused, the computer screen is on, and there has been mouse or keyboard activity on the tab within the past minute. Time on site per day is measured as the aggregated time across sessions from midnight to midnight in the participant's local timezone. There was one data point per user per day per targeted web site. Because time data is not normally distributed, we adopt a common practice of log-transforming the time data prior to analysis.

We measured *attrition* as the number of days the participant kept the extension enabled. We also noted if the extension was still enabled at the conclusion of our study. The browser does not send a notification to our server if a user disables the extension, so we coded instances of attrition when the server stopped receiving data from the user for over two days, with no later resumption.

As with many online field experiments, effective data cleaning is essential to accurate analysis. We excluded users who had HabitLab installed on multiple devices, to focus on site usage on a single device. We discarded days on which the target site was never visited, as in neither condition would the intervention have been shown. We also discarded the first day because participants installed the extension midway through the day, resulting in an underestimate at the day level; we likewise discarded any days on which the user uninstalled or disabled the extension, as this would again cause the measured time to be an underestimate of the actual time spent on site that day.

4.4.5 Method of Analysis

For analyzing effectiveness at both the day and session level, we used a linear mixed model (LMM). We used an LMM because we have multiple samples from each user, but the number of samples from each user and in each condition is variable (because attrition may occur before they completed all conditions, or they may not visit a site on a particular day), which violates the assumptions of the repeated-measures ANOVA.

To test whether interventions decrease in effectiveness over time (H1), we focused on just data points from the static condition. The model included a term for the number of days that particular intervention had previously been seen,[1] a random effect for the participant, and a random effect for domain. To test linear mixed models for significance, we used a likelihood ratio test to compare a reduced model without the number of days predictor to the full model. A significant test indicates that the number of days has statistically significant explanatory power, analogous to a significant beta coefficient in a traditional regression.

To test whether static or rotated interventions increase effectiveness (H2), we used data from both the static and rotation conditions. This second LMM, predicting log time spent on the site each day, included a random effect for participant, a random effect for domain, a fixed effect for block length, and a fixed effect for condition. A likelihood ratio test compared to a reduced model without the condition variable.

To analyze whether static or rotated interventions increase attrition (H3), we used a Cox proportional hazards regression. Cox proportional hazards models predict the relative "hazard" (i.e. risk) of attrition given each predictor. This is used in the health sciences for estimating expected lifetimes when we may have differing durations of observations for each participant, and may have observed deaths (which correspond to attrition) for some participants but not others. Each data point consists of a point of observation, and whether the participant had experienced attrition at that point or was still active. To avoid crossing conditions in this analysis, we focus the Cox analysis on just each user's first assigned block and condition, for example a seven-day rotation block or three-day static block. Each observation consists of the length of block, and whether the user had experienced attrition by the end of the first condition for their first block. The Cox model used a single predictor: condition. The output of a Cox proportional hazards model is similar to a regression, with a significance value and estimate attached to the predictor.

[1] Repeating the analysis using the number of times the intervention had been seen yields the same conclusions.

Table 4.1: Within the static condition, interventions decline in effectiveness. Longer visit lengths increase with the number of days seeing the same static intervention.

	Dependent variable:
	Log time spent per day
Number of days the user had seen the static intervention	0.225^{*}
	(0.097)
(Intercept)	4.759^{***}
	(0.392)
Observations	124
Note:	*p<0.05; **p<0.01; ***p<0.001

4.4.6 Results

In this study, participants had an average of 3.0 target sites enabled. They visited at least one target site 67% of days on average. On each of those days, participants experienced interventions an average of 3.6 times. We did not receive any feedback indicating that participants were aware of patterns in how HabitLab was rotating interventions.

Effectiveness of interventions over time

First we examine whether interventions decrease in effectiveness over time within the static condition. If so, rotation may be a viable strategy.

The likelihood ratio test confirms that the number of days the user had seen the static intervention affected the log of time spent on a domain per day ($\chi^2(1) = 4.69, p < 0.05$), supporting H1. Each day the intervention has been previously seen increased the log time spent by 0.225 (Table 4.1). By exponentiating the log estimates, this translates into an increase of 25% on top of a baseline 117 seconds per day for each additional day the user were exposed to the static intervention.

Effectiveness of rotation and static intervention strategies

Next, we compare whether the daily time spent on domains differs between days when participants were in the rotation and static conditions.

The likelihood ratio test found a significance difference between the full and reduced models predicting effectiveness ($\chi^2(1) = 4.88, p < 0.01$), indicating that condition significantly impacted effectiveness. Relative to the static condition, rotating interventions decreased the log time spent on domains per day by 0.417 (Table 4.2), supporting H2. Exponentiating the coefficients for descriptive purposes, this translates into a shift from an estimated 146 seconds per day in the static condition to 96 seconds per day in the rotation condition, a decrease of 50 seconds (34%) per day.

Table 4.2: Daily time spent on sites in the static and rotation conditions. Users spend less time per day on sites in the rotation condition.

	Dependent variable:
	Log time spent per day
Rotation (baseline: static)	-0.417^{*}
	(0.190)
Block length	0.018
	(0.048)
(Intercept)	4.981^{***}
	(0.346)
Observations	370
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 4.3: A Cox proportional hazards analysis suggests that the rotation condition substantially increases the hazard of attrition. Coefficients are log hazard ratio, so positive values indicate increased hazard and negative values indicate decreased hazard.

	Dependent variable:
	Log hazard ratio
Rotation (baseline: static)	0.544^{*}
	(0.249)
Observations	217
Note:	*p<0.05; **p<0.01; ***p<0.001

Attrition due to rotation and static intervention strategies

The Cox proportional hazard regression model comparing the static and rotation conditions found that attrition rates are significantly higher with the rotation condition (Figure 4.2, Table 4.3). After 2 days, 78% of users remain in the static condition, while only 71% remain in the rotation condition. After 7 days – the duration of the longest experiment block – 68% of users remain in the static condition, while only 39% of users remain in the rotation condition. These results support H3.

We considered the possibility that switching between static and rotated interventions contributes to attrition beyond simply rotating them. We analyzed this by comparing the probability of attrition on days where the condition remains the same as the previous day, to days where the condition changes – either from static to rotated, or from rotated to static. The baseline daily attrition rate is 18% when staying within the same experimental condition – 14% when staying within the static condition, and 20% when staying within the rotation condition. On the first day after switching from static interventions to rotated interventions, the attrition rate is 36% – a significant increase compared to remaining within the same condition (Fisher's exact

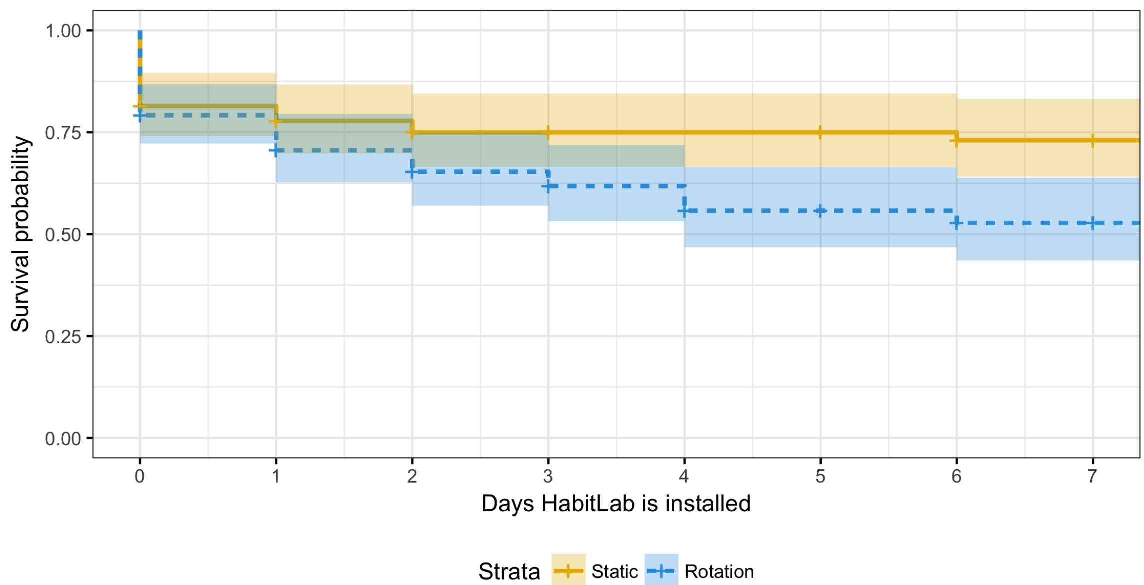


Figure 4.2: Rotating interventions increases attrition among users.

test, $p < 0.001$). However, switching from rotated interventions to static interventions does not increase the attrition rate – it remains at 18%. So we believe these effects are not due to the changes between conditions, but due to the conditions themselves — switching from static to rotated is experiencing the first instance of a rotation, and it is not surprising that the effect may be larger with the first change.

4.5 Study 2: Longer-Term Effects of Rotation on Attrition

Study 1 found that compared to static interventions, rotation increases effectiveness but also increases attrition. To provide additional support for our findings in Study 1 and motivate our design experiment, we present a second field study that seeks to answer the question: Does the number of interventions in the rotation affect the level of attrition? This study occurs over a longer period — ten weeks — allowing us to examine these effects in a more longitudinal setting.

4.5.1 Participants

Our participants were HabitLab users who installed over a 5 week period in January–February 2018 and consented to our experiment protocol. 680 users who agreed to participate. After excluding users with multiple devices, users who did not complete the onboarding process, and users who had less than two sessions on Facebook where they saw interventions — we restricted analysis in this study to users who were using Facebook because it had the most number of default interventions available — we were left with

409 participants. Demographics were similar to Study 1.

4.5.2 Method

This was a between-subjects study where users' default settings for the number of enabled interventions varied depending on their condition: some users only had one default enabled intervention, and others had more. Interventions were then selected randomly from the enabled set. Among users who did not change these defaults, this enabled a between-subjects comparison of the effects of the number of interventions a user was rotating between, on retention rates.

In practice, we found that many users changed the set of interventions — 78% of participants in this study changed them over the course of using HabitLab, most often during onboarding. We wanted to retain a good user experience, but this muddied the experimental manipulation. So, we restricted analysis to the 91 users who did not change defaults. A χ^2 test found there was no significant effect of condition on whether users changed defaults ($\chi^2(2)$= 0.4671, p=0.8), suggesting that randomization remained effective even after this filter.

Unlike Study 1, this was a between subjects experiment, so there were no time blocks: participants were assigned to the condition for the duration of the study.

4.5.3 Conditions

Participants were randomized into three conditions. In the one intervention condition, for every site the user enabled HabitLab on, only one intervention was enabled by default. The intervention was randomly chosen among the set of default interventions for that site. This is equivalent to the static condition from Study 1.

In the all interventions condition, for every site the user enabled, all interventions that are default for that site were enabled by default. This is equivalent to the rotation condition from Study 1. In the half interventions condition, for every site the user enabled, half of all interventions that are default for that site were enabled by default. The subset was chosen randomly.

4.5.4 Measures

We measured attrition, using the same procedures as those described in Study 1.

4.5.5 Method of Analysis

Like Study 1, we applied a Cox proportional hazards regression model to compare attrition rates.

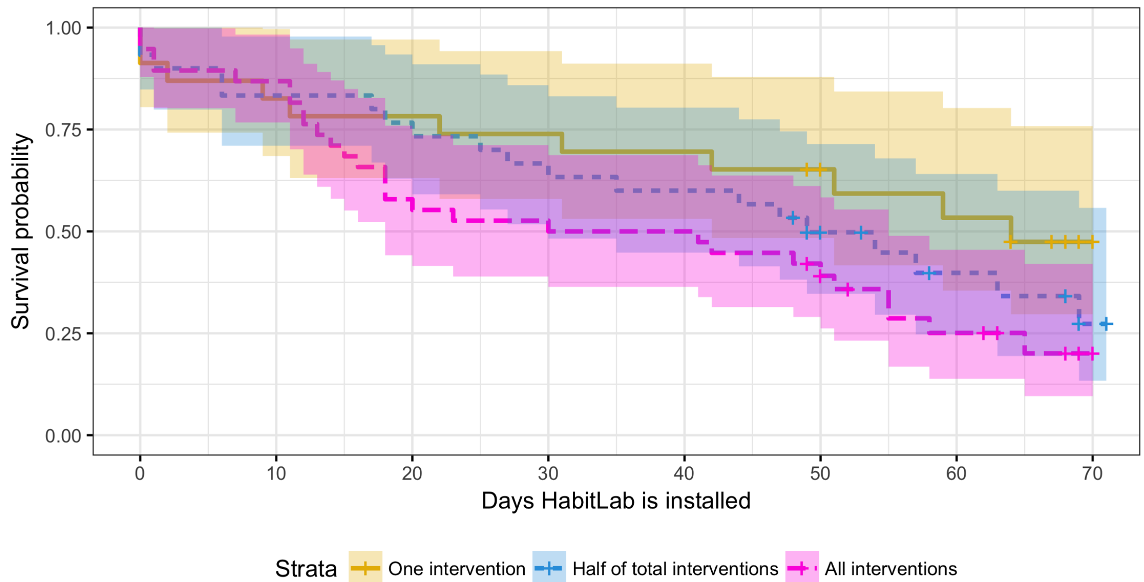


Figure 4.3: Including all interventions resulted in significantly more attrition than just one intervention.

4.5.6 Results

In this study, participants had an average of 3.3 target sites enabled. They visited at least one target site 64% of days on average. On each of those days, participants experienced interventions an average of 6.8 times.

In this longer, between-subjects experiment, attrition rates were significantly higher in the all interventions condition (Figure 4.3, Table 4.4). This agrees with the analogous result from Study 1 showing a higher attrition rate for the rotation condition. The half of total interventions survival curve falls in between that of the one intervention and all interventions conditions, but does not have a statistically significant difference.

RQ4: Are users who enable or disable interventions during onboarding less likely to attrition?

Table 4.4: A Cox proportional hazards analysis over a longer period suggests that rotating with more interventions increases the hazard of attrition.

	Dependent variable:
	Log hazard ratio
Half of total interventions (baseline: one intervention)	0.395
	(0.380)
All interventions	0.711^{*}
	(0.358)
Observations	91
Note:	*p<0.05; **p<0.01; ***p<0.001

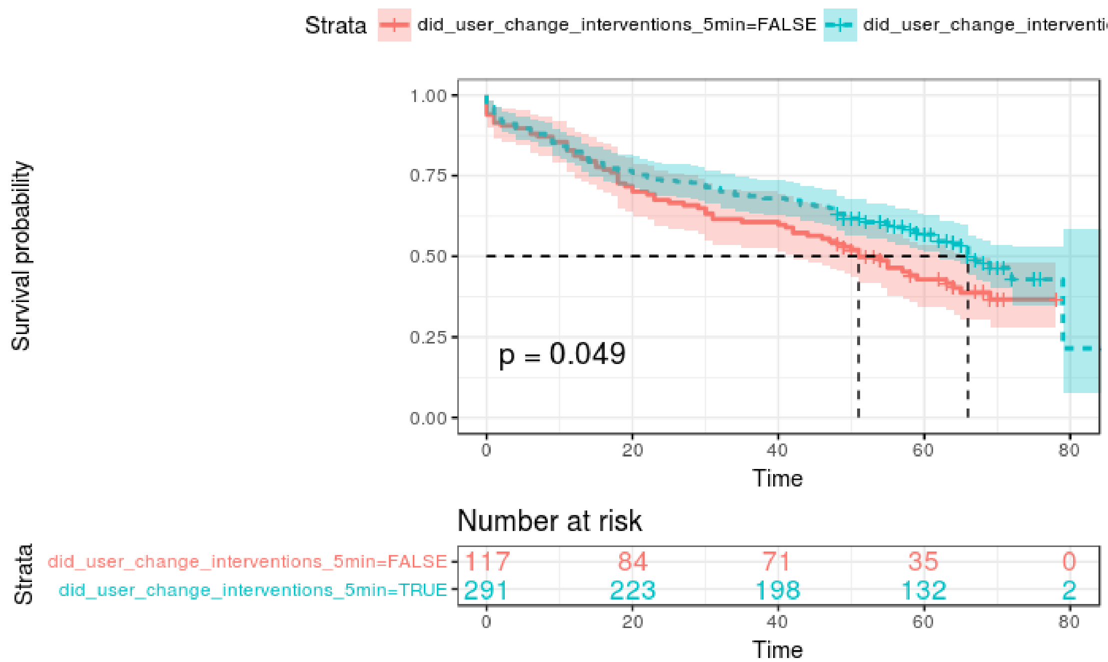


Figure 4.4: Users who change interventions during the first 5 minutes will be significantly less likely to attrition later.

We compared users who enabled or disabled interventions during the first 5 minutes, compared to other users who completed the onboarding process without enabling or disabling any interventions. We found that users who enabled or disabled interventions during the first 5 minutes had a significantly lower rate of attrition, as shown in Figure 4.4.

4.6 Study 3: Design Interventions to Reduce Attrition

Study 1 and Study 2 collectively demonstrated that rotation increases effectiveness but also increases attrition. Why does rotation increase attrition? To understand this, we needed to understand why users uninstalled in the first place.

We performed a qualitative content analysis on the uninstall feedback left to us by users. This feedback was collected in a tab that opened automatically when users uninstalled HabitLab. The page stated that feedback would be used for research purposes. Users had the option to check boxes to agree with a set of predefined reasons they why they were uninstalling, and leave free-text feedback. We performed an inductive analysis of the free-text feedback, grouping responses by themes, reflecting on our themes, and refining our groupings until convergence.

A total of 782 users submitted the uninstall feedback form. This data represents all past users of HabitLab, and includes users outside studies 1 and 2. We use this larger dataset because only 8 participants from Study 1, and 39 from Study 2, filled out the feedback form. 751 users who submitted the form checked at least one of our predefined reasons. 274 users (36%) uninstalled because "Interventions were annoying", 248 users (33%) uninstalled because HabitLab "Did not feel effective", 100 users (13%) uninstalled because HabitLab "Was causing lag", 75 (10%) uninstalled due to "Privacy concerns", and 202 (27%) cited "Other reasons". The total sums to more than 100% because users could check more than one reason.

A total 155 users submitted free-form textual feedback. Some users began with an incorrect mental model and uninstalled after they learned what it was doing:

- *Didn't seem what I was expected. Installed two minutes ago and removed it*
- *I just didn't understand the concept before downloading*

Some users indicated they wanted more control over the intervention that was shown to them, or were simply looking for a time tracker and were not interested in interventions at all:

- *I wanted a timer for every "domain", it can be good for statistics of time*
- *I was interested in tracking my usage to start, instead of setting interventions that I may not actually be concerned about*

Some users indicated dissatisfaction with particular interventions:

- *Mostly it was the bar covering up facebook message indicators*
- *it was just annoying you out of not using sites, not convincing you to. It became like ads, they are always there. But you don't like them and turn them off with ad-block.*

Some users wished interventions would be more forceful, or less intrusive:

- *Interventions are not forceful enough. They are too easy to click around or disable*
- *I liked the interventions but not on every page change or load, that was just a bit too much*

Finally, some users decided they simply did not want or need interventions:

- *Made me realize I don't have Facebook addiction, spending less than 30 minutes [...] per day*
- *I'm weak...*

Other themes included localization issues, performance issues, privacy concerns, accidental installations, and misattribution of other issues to HabitLab.

ome feedback indicated that users had an incorrect mental model of the system: they were expecting to see a particular intervention consistently, instead of having them rotate:

does not fucking work, hiding the newsfeed in particular

Other users began with an incorrect mental model and uninstalled after they learned what it was doing:

Didn't seem what I was expected. Installed two minutes ago and removed it

I just didn't understand the concept before downloading and it's intentions aren't my demons as it happens

Some users indicated they wanted more control over the intervention that was shown to them, or were simply looking for a time tracker and were not interested in interventions at all:

I wanted a timer for every "domain", it can be good for statistics of time

I was interested in tracking my usage to start, instead of setting interventions that I may not actually be *concerned about*

Would have preferred a gentle log - perhaps emailed - giving usage statistics. In present form this operates like pop-up ads. Still, it was reasonably insightful into understanding my usage for the time I used it

Some users indicated dissatisfaction with particular interventions:

Mostly it was the bar covering up facebook message indicators

You covered up useful buttons. Don't do that

Made Facebook unusable. Which might be the point?

Some users wished interventions would be more forceful:

Interventions are not forceful enough. They are too easy to click around or disable

Some users were getting fatigued from seeing too many interventions:

Looked like some nice options for streamlining sites, but it's actually a nanny. I don't need a nanny whining at me.

I liked the interventions but not on every page change or load, that was just a bit too much

it was just annoying you out of not using sites, not convincing you to. It became like ads, they are always there. But you don't like them and turn them off with ad-block.

Some users disliked that the intervention was reminding them that they were visiting sites:

I noticed that I was going on imgur, youtube, facebook (my choice of addictive sites) more, after I had installed the extension. So, I'm uninstalling. I think the extension made me more conscious of the fact that I was visitng the sites, but maybe the rewards were making me go back to the site? I'm not sure

There were also users who cited localization issues, or may not have understood the text presented during onboarding because the app was not localized to their native language:

Translate in french please – This was from before we localized to French

non lo voglio – Italian for I don't want it. The extension is not localized to Italian

Some users cited performance issues, bugs, or conflicts with other extensions:

Was awesome, but was making chrome really slow, i mean really slow! Seems like you need to fix some memory issues

It's quite possible something else was causing lag – but lag was there. I also was just checking it out. I don't really use facebook or youtube

Catastrophic stability problems after installation; may be due to a different extension

Seeing if this extension is causing gmail compatibility issues

Wasn't sure if it is effecting battery life

Some users had simply been testing the app or evaluating alternatives:

I love the app - I'm just removing temporarily to see if it's affecting another app (Freedom.to)

I want to try other types of chrome extensions to block time-consuming websites and don't want to mess with your data

I prefer the "Forest" application

I forgot that I already had another program that did essentially the same thing for the computer in general instead of just for this particular browser

Some users had privacy concerns:

I was just worried, I mean it's (Anonymized) and all, but a bunch of students tracking everything I do and all my browser history, that just felt too much of a price to pay

Finally, many uninstalls were simply due to the extension being automatically installed on a non-work computer due to Chrome's behavior of automatically installing extensions across all devices, which is why we had restricted analyses to just users who used one device:

Neighbor installed this to my computer without my consent!

Someone else installed. Did not want it

Dont need it on work computer, just at home

this is my wasting time computer and i dont need it

Some users decided they simply did not want or need interventions:

Made me realize I don't have Facebook addiction, spending less than 30 minutes of my desktop time on it per day

I rarely waste time on my desktop. This would be much more useful on my mobile phone

I just don't use my laptop as much as I thought would be necessary for an intervention

Some users uninstalled due to misattributing other issues to our software:

For some reasons Facebook blocked me and I am trying to figure out the reasons. What I know that the Habitlab extension was deactivated (not by me) and then, boom, I was blocked

Finally, some users reached their goals and decided they simply didn't need the extension anymore:

I reached my goal to reduce time spent on certain pages. Thanks folks!

I used the extension to curb my Facebook habit and eventually gave up Facebook altogether - something I've been wanting to do for a long time. Thank you

4.6.1 Design interventions

Based on the qualitative feedback on reasons for uninstalling, we drew on two of the most consistent themes to hypothesize why rotation may be increasing attrition:

Hypothesis 4 (H4)**.** *Violation of mental model: Users may have sped through onboarding and not understood that HabitLab rotates interventions. So, when they experience a new intervention, the system violates their mental model and they disable it in confusion or frustration.*

Hypothesis 5 (H5)**.** *User control: Users may be aware that the system is choosing interventions for them, but are frustrated by a lack of control over the system's behavior. They may dislike one or more of the interventions but not realize how to turn them off.*

These two hypotheses could feasibly be addressed through design interventions. blackOther pieces of feedback, for example how aggressive the interventions were, we judged as out of scope of the current study

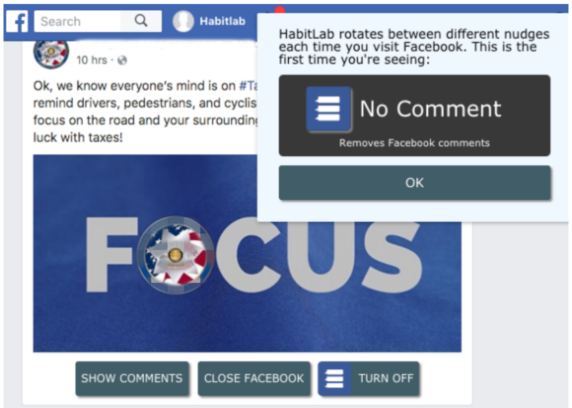


Figure 4.5: Mental model interface: each time the user sees a new intervention, HabitLab names it and explains about rotation.

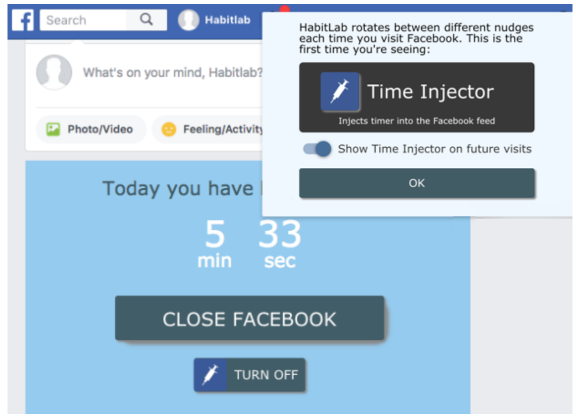


Figure 4.6: User control interface: in addition to the mental model information, HabitLab gives users a direct interface to disable the new intervention.

on rotation strategies and will pursue as future work. We developed two different interfaces, one to address mental model violation and the other to address a perceived lack of control. They are shown to users when they see a new intervention for the first time.

The first design, which we will call *mental model* (Figure 4.5), is inspired by H4: it reminds the user that HabitLab has rotated to a new intervention and gives the name of the intervention. If mental model misalignment was the issue, this design might help explain to the user what the system is doing and why. The second design, which we will call *user control* (Figure 4.6), is inspired by H5: it includes the message in the information design but also adds a toggle option to allow the user to turn off the new intervention for future visits without needing to visit HabitLab's settings. If lack of control was the issue, this design may give sufficient control so that users keep HabitLab enabled.

e hypothesized:

Hypothesis 6 (H6). *The information and control designs will have less attrition than a control design.*

4.6.2 Experiment Design

We ran a between-subjects design where we randomized the design shown to new users of HabitLab and tested whether it impacted attrition over a period of one week, similar to Study 1.

4.6.3 Participants

Our participants were HabitLab users who installed over a 10 day period in April 2018. There were a total of 282 users who installed and agreed to participate. We removed users who were not new users (e.g. an existing user installing on a new device, or a former user reinstalling the system), and users who left before they saw their first intervention. This leaves us with data from 93 participants. Demographics, estimated by Google Analytics, were similar to Study 1.

4.6.4 Method

Participants installed HabitLab and set it up as described in the Study 1 and Study 2. They used HabitLab in the course of their normal web browsing activity. HabitLab rotated between randomly chosen interventions on each visit to the chosen web page for all users, equivalent to the rotation condition in Study 1. Each time the user experienced a new intervention that they had not seen before, however, HabitLab might show an explanation design in the corner of the browser.

4.6.5 Conditions

There were three conditions for this study. In the no design condition, users saw no message, equivalent to the rotation condition from Study 1. In the mental model condition, users were shown the informational intervention (Figure 4.5) to remind them that the system rotates interventions. In the user control condition, users were additionally given control over whether to turn off each new intervention without needing to visit the settings screen (Figure 4.6).

4.6.6 Measures

Our main dependent variable was attrition—how many days users kept the system installed by the end of the study, seven days after installation. The measure of attrition was the same as in Study 1 and Study 2. We also measured effectiveness, using the same method as Study 1.

4.6.7 Method of Analysis

To analyze attrition, we again used a Cox proportional hazards regression model, similar to Study 1, using interaction design as the predictor variable. To analyze effectiveness, we used a LMM predicting log time on site per day, with a fixed effect for condition, and random effects for participant and domain. Data cleaning followed the same procedures as Study 1.

4.6.8 Results

In this study, participants had an average of 2.9 target sites enabled. They visited at least one target site 71% of days on average. On each of those days, participants experienced interventions an average of 6.6 times.

The Cox proportional hazard regression indicates that the mental model design significantly reduces attrition rates relative to no design ($p < 0.05$, Figure 4.7, Table 4.5). This result supports H4. After seven days, 79% of participants in the mental model condition remain, while 80% remain in the user control condition and only 44% remain in the no design (control) condition. In other words, the intervention coditions more

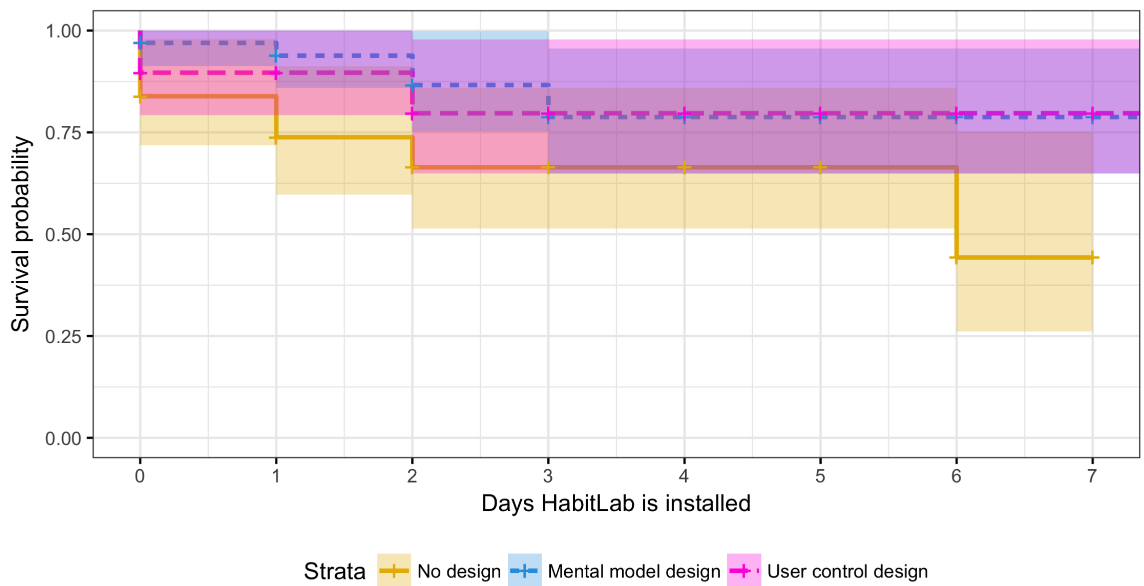


Figure 4.7: Reminding users about how the rotations worked every time a new intervention was introduced significantly reduced attrition rates.

than halved the attrition rate, from 56% to 21% attrition. Adding the additional option to permanently turn off the intervention the first time it is seen is not significantly different from no design given our sample size.

There was no effect of condition on effectiveness: the full model was not significantly more explanatory than the reduced model without the condition variable ($\chi^2(2) = 1.46, n.s.$). So, these interventions did not reduce effectiveness while they were improving attrition.

4.7 Discussion

Our findings suggest that changing behavioral interventions can be beneficial from the perspective of efficacy, but detrimental to retention. By showing simple messages when presenting new interventions, we can improve users' mental models, and reduce attrition from changing interventions.

In addition to the interface-based techniques we have presented to combat detrimental effects of changing interventions, algorithmic techniques can also help. For example, in the context of a multi-armed bandit, potential algorithmic techniques include:

1. Limiting the exploration speed such that users are not overwhelmed by the rate at which they are seeing new interventions.

2. Modeling individual interventions' likelihood of attrition, and favoring algorithms which are less

Table 4.5: A Cox proportional hazards analysis suggests that the informational intervention that corrected users' mental models was successful in reducing attrition due to rotation. Coefficients are log hazard ratio.

	Dependent variable:
	Log hazard ratio
Mental model design	-1.015^{*}
	(0.494)
User control design	-0.869
	(0.527)
Observations	93
Note:	*p<0.05; **p<0.01; ***p<0.001

likely to cause attrition if needed to keep the user around longer.

There are also additional interface-based techniques that may be helpful in reducing attrition from changing interventions, but that we have not tested, such as:

1. Making how new interventions are introduced predictable and known to the user.
2. Allowing users a choice of intervention when we introduce new interventions.

4.7.1 Limitations

This work featured deployment periods of a few weeks. This may not be enough time to observe some very long-term effects: for example, some changes in intervention effectiveness set in only after months or years [76]. That said, given the fast turnover rate which is observed with behavior-change software, even short-term effects of changing interventions on attrition can be important.

While we believe our general finding about the double-edged nature of changing interventions may apply to other behavior-change contexts, particular parameters—such as speed at which users grow blind to an intervention, may be domain-specific.

One shortcoming of our Study 1 design is that we cannot rule out the possibility that our observed increase in effectiveness is due to selective attrition, rather than due to benefits from the rotation. Namely, it is possible that observing rotation may selectively lead to uninstallation for users for whom interventions are ineffective. To rule out this possibility, we will need to investigate ways to maintain retention in the presence of rotation, and see whether the improvement in effectiveness relative to a static intervention still remains. It may also be possible to design intention-to-treat analyses that discount attrition in measures of effectiveness. blackFurthermore, while we observed that the first visit is longer than subsequent visits when users visit sites multiple times per day, but this effect may be due to temporal usage patterns rather than intervention effectiveness.

blackBecause users have differing preferences, interventions may have differing rates of attrition for each user. An ideal retention-maximizing system would not assign interventions randomly, but would personalize interventions to each user. Assuming there is a novelty component to attrition — i.e., users quit because they grow bored of the same intervention — then a system which intelligently times interventions to minimize attrition can in theory have lower attrition than even the best static intervention. There are 2 difficulties in making this a reality: first is needing to learn to correctly predict which intervention would minimize attrition for a user at a given time, a reinforcement learning problem. Second, as shown by the increase in attrition when using a naïve rotation strategy, a system that switches between interventions also needs to overcome the barriers of needing users to develop more complex mental models, and ensuring that users feel in control.

4.7.2 Design reflections on social computing and behavior change

Social systems are inherently tied to behavior change and retention. Social networks and other social apps and services make heavy use of gamification and behavior change techniques to drive engagement and boost retention [45, 25]. A system like HabitLab that helps users use these services less thus occupies an interesting space: it is modifying the service to hide the features that serve to boost engagement, helping users break away from their addiction to the site.

But we tread a fine line: behavior change systems themselves suffer from attrition, so we may sometimes need to make tradeoffs between better retaining users and helping them regulate their behaviors. For example, the Facebook interventions in HabitLab with the lowest attrition—those that passively show time spent—are among the least effective. Is telling users that the system is helping them more than it actually is a form of benevolent deception [4] that would ultimately help boost retention and help users achieve their goals? Would gamifying the system with social features, making users connect with friends and keep tabs on their friends' social media usage, help boost retention and effectiveness—even though users may lose time engaging with social features?

4.8 Conclusion

Behavior change intervention effectiveness declines as interventions are repeatedly shown to the user. This decline can be combated by rotating between a stable set of different interventions. Rotating interventions increases attrition, but user interface changes can ameliorate the issue. Taken together, these results suggest opportunities to build behavior change systems that operate more like coaches and tutors: they might explore different strategies to find what works well, and then occasionally rotate to keep things fresh.

Table 4.6: Within the static condition, interventions decline in effectiveness, with longer visit lengths with increasing larger number of days since it was first observed.

	Dependent variable:
	Log time spent per session
Number of days the intervention has been seen	0.056^{***}
	(0.021)
(Intercept)	3.826^{***}
	(0.143)
Observations	1,007
Note:	*p<0.1; **p<0.05; ***p<0.01

4.9 Appendix: Replication of Study 1 Findings using Session Level Measurements

This appendix replicates the findings of Study 1 using an alternative method of analysis, looking at time on site per session rather than day.

Time on site per session is measured as the total time the user was actively using a site in a browser tab, from when they visited the site until they closed the tab. If the user switches tabs to a different site, the time spent on the other site is not counted towards the current session time.

To determine whether the user is actively using a target site, we use Chrome's internal definition of active – the browser window and tab is focused, the computer screen is on, and there has been mouse or keyboard activity on the tab within the past minute. Because time data is not normally distributed, we adopt a common practice of log-transforming the time data prior to analysis.

4.9.1 Effectiveness of interventions over time

The likelihood ratio test confirms that the number of times a user has seen an intervention affected the log of time spent on a domain per session ($\chi^2(1) = 6.69, p < 0.01$), supporting H1. Each time the intervention has been previously seen increased the log time spent by 0.05633 (Table 4.6). By exponentiating the log estimates, this translates into an increase of 5.8% on top of a baseline 46 seconds per session for each additional time the user saw the intervention during the study.

An alternative method of analysis, where we measure the raw number of times the intervention has been seen instead of the number of days it has been seen, yields the same results. Restricting analysis to just Facebook also yields the same results.

4.10 Appendix: List of interventions used in this study

The following is the list of interventions used for this study, showing the intervention name and description as seen by the end user. As this study was run before the mobile version of HabitLab existed, these interventions are all for the browser version.

There are 27 interventions total: 7 generic interventions that can be used on all sites, 5 interventions designed specifically for Facebook, and additional ones designed specifically for YouTube, Reddit, Twitter, Netflix, Gmail, Amazon, iQiyi, and Youku

Generic interventions that can be used on all sites:

- Minute Watch: Notifies you of time spent every minute
- Supervisor: Shows time spent on site at the top of screen
- Scroll Freezer: Freezes scrolling after a certain amount of scrolls
- Stat Whiz: Show time spent and visit count each visit
- GateKeeper: Makes you wait a few seconds before visiting
- 1Min Assassin: Closes tab after 60 seconds
- Bouncer: Asks how long you want to spend on site this visit

Facebook-specific interventions:

- Time Injector: Injects timer into the Facebook feed
- Feed Eater: Removes the Facebook news feed
- TimeKeeper: Notifies you of time spent in the corner of your desktop
- No Comment: Removes Facebook comments
- Clickbait Mosaic: Removes clickbait from the news feed

Youtube-specific interventions:

- Sidekicker: Remove sidebar links
- Think Twice: Prompt the user before watching a video
- No Comment: Removes comment section

Netflix-specific interventions:

- Fun Facts: Gives you a fact and links an article on the effect of TV
- Alarm Clock: Asks the user to set an alarm before watching a show
- Stop Autoplay: Stops the site from automatically playing the next video

Reddit-specific interventions:

- Comment Remover: Removes Reddit comments
- Mission Objective: Asks what you aim to do this visit and puts a reminder up

Youku-specific interventions

- Think Twice: Prompt the user before watching a video
- Sidekicker: Remove sidebar links

iQiyi-specific interventions

- Think Twice: Prompt the user before watching a video
- Sidekicker: Remove sidebar links

Twitter-specific interventions:

- Feed Eater: Removes the Twitter news feed

Amazon-specific interventions:

- No Recs: Hides recommendations

Gmail-specific interventions

- Speedbump: Delays the arrival of new emails

Chapter 5

Do Productivity Interventions Save Time or Just Redistribute It?

5.1 Introduction

We use productivity behavior change interventions to try to keep ourselves in focus. But do these systems truly save us time? Or do they just redistribute the time elsewhere? In other behavior change domains, interventions sometimes have effects on behaviors other than the ones they were targeting [122, 33].

One possibility is that interventions narrowly impact just the goal that they target, and have no effect on time spent elsewhere. We will refer to this as the *isolated effects hypothesis*. Taking the relationship between time spent on Facebook and Instagram as an example, the isolated effects hypothesis would predict that an intervention that helps reduce time on Facebook should have no effect on time spent on Instagram. Persuasive systems often claim to result in the intended behavioral changes without observable consequences elsewhere, lending support for this hypothesis [14, 34, 83, 9, 10]. If the isolated effects hypothesis is true, overall productivity can be boosted through interventions that individually target each goal.

However, people have a limited supply of willpower [16], can maintain focus for only so long [64, 36, 87], and need downtime — so perhaps the time saved is actually just redistributed to other unproductive applications. We will refer to this as the *redistribution hypothesis*: saving time on one unproductive application results in an increase in time spent on other unproductive applications. Returning to our example of a productivity intervention targeting Facebook, redistribution would hypothesize that an intervention that reduces time on Facebook will increase time spent on Instagram. Redistribution may be partial, where the time redistributed is some fraction of what was saved. Or more bleakly, redistribution may be total, where the time redistributed is entirely shifted to other applications and there is no overall improvement in productivity.

A third possibility is that saving time on one application breaks a habit loop [41] and reduces time

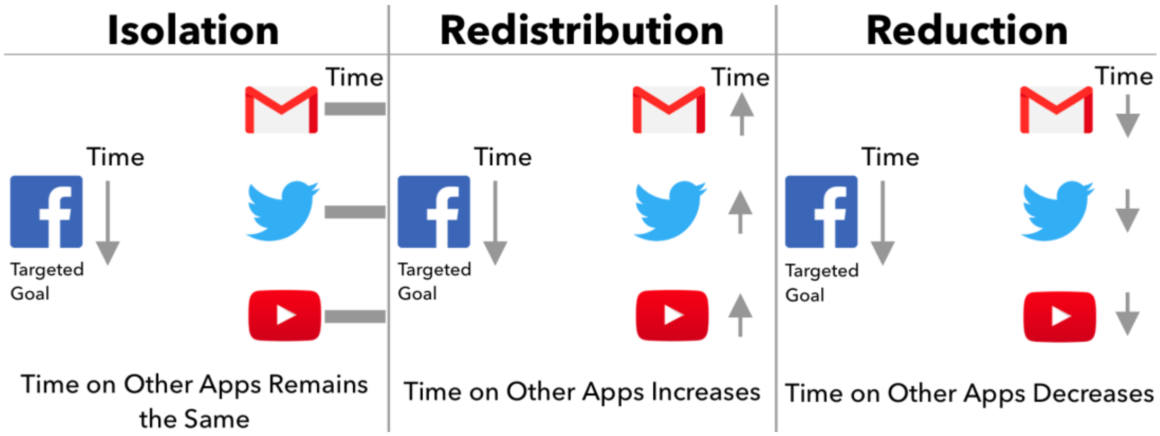


Figure 5.1: When interventions reduce time on a targeted goal such as Facebook, the time saved may (left) be isolated from effects on other goals, (center) be redistributed to other goals, or (right) decrease time spent on other goals.

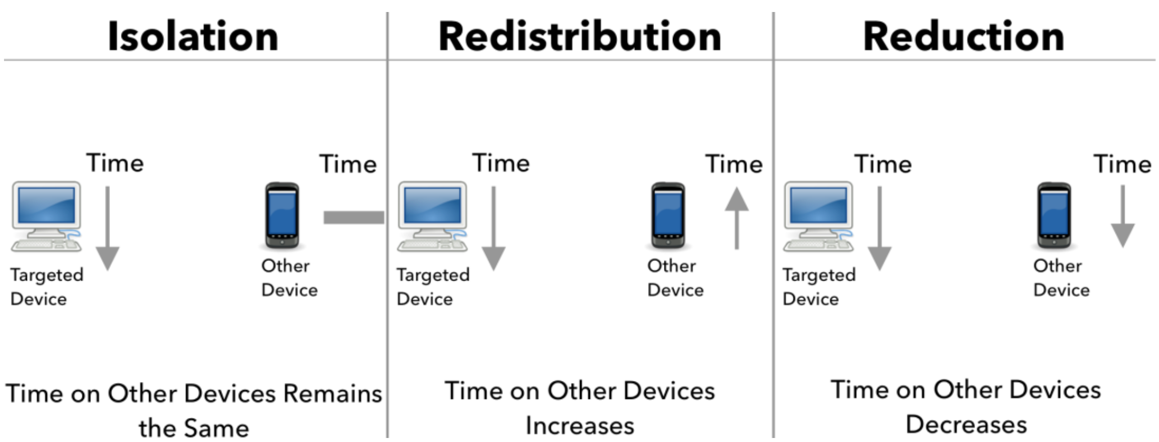


Figure 5.2: When interventions reduce time on a targeted device e.g. a browser, the time saved may (left) be isolated from effects on other devices, (center) be redistributed to other devices, or (right) decrease time spent on other devices.

spent on other applications as well, so the actual net improvement in productivity is even better than just what is saved on the target application. We will refer to this as the *reduction hypothesis.* Returning to our example of a productivity intervention targeting Facebook, this would hypothesize that an intervention that reduces time on Facebook will also reduce time on Instagram. Perhaps once we enter "procrastination mode" and visit one unproductive application, we wind up chaining together visits to another unproductive application, and another—but if a productivity intervention helps us break the chain early on, we will never visit the later unproductive applications.

These three hypotheses lay out the three possibilities of what happens to other goals when we intervene on a focal goal (Figures 5.1–5.2): time on those other goals might stay the same (isolated effects), go up (redistribution), or go down (reduction). In this paper, we seek to adjudicate between these hypotheses using HabitLab [75], an in-the-wild productivity experimentation environment that users can voluntarily participate in by installing. Prior work described HabitLab as a Chrome browser extension; in this paper we created and deployed a companion HabitLab Android application, allowing us to study any redistribution of time that might be happening across devices, as when a user avoids Facebook on their browser but ends up checking Facebook on their phone instead.

After installing and agreeing to our experimental protocol, users specify what they wish to reduce time on, which we term *goals.* In the case of the Android version, goals take the form of applications (apps), whereas on the Chrome extension goals are sites. We then deploy interventions to help users reduce their time on these goals, which can appear when the user visits a website (Chrome) or app (Android). To study redistribution, we periodically manipulate the frequency at which interventions appear for each goal — if the goal is in the frequent condition that week, it will appear every time the user visits that application, whereas if the goal is in the infrequent condition that week, it will appear on 20% of visits. This experimental design allows us to observe the effects of a goal being in the *frequent* setting not only on how much time users spend on that focal goal, but also what happens to time on other goals when that focal goal is in the frequent setting.

Our analysis first begins by seeing whether interventions are effective at reducing time on the focal goal, disregarding any possible redistribution effects. We do so by comparing time spent per day on the application on weeks where interventions are shown frequently, vs those weeks where interventions are shown infrequently. We find that they are effective, with time spent on goal sites reduced by 8.0% on the Chrome version, and time spent on goal apps reduced by 37.3% on the Android version.

Next, we investigate whether time is redistributed to other sites/apps on the same platform (browser or mobile) when interventions are frequently shown. We find that giving interventions within the browser produces a reduction effect, with users using sites/apps less when there are more interventions shown on other sites/apps – however, effects of interventions are isolated on mobile.

Finally, we investigate whether time is redistributed across devices. We do not observe any significant time redistribution effects in either direction.

This paper contributes a look into potential unintended side effects of productivity interventions on other sites, apps, and devices. We find that productivity interventions do not appear to have deleterious second-order effects on goals other than the ones they are targeting, and in some cases, may even have beneficial second-order effects by breaking habit loops.

5.2 Research Questions

Unproductive time begets further unproductive time. For example, increased time spent online can increase sleep debt, which in turn leads to more time spent online [88]. Likewise, the Hook Model claims that many of the most addictive online sites use a cycle of investment techniques to keep users coming back—for example, making a post on Facebook may result in future notifications, which will in turn will get the user to come back and make more posts [45]. Finally, sites such as Facebook, Reddit, Twitter, and Buzzfeed are filled with links to each others' content, so it may be the case that increasing usage of one will increase usage of others. If productivity interventions are able to break this vicious cycle of procrastination for one application, they may actually reduce time spent on other unproductive applications as well.

The importance of understanding the effectiveness of productivity interventions in a complete ecosystem and the rising awareness of unproductive time spent on mobile devices call into focus: would productivity interventions reduce net unproductive time? Or is it a weak palliative with little discernible effect? This led to our research question:

Research Question (RQ)**.** *Do productivity interventions reduce net unproductive time, or just redistribute it to other applications, sites, and devices?*

Studying behavior change effects across multiple devices is important: focusing on a single platform will myopically miss unproductive behaviors on other platforms. Attention is fragmented in both mobile and traditional desktop environments [78, 85]. The time spent on mobile devices has increased more rapidly than time on computers or TVs [17, 30]. On the other hand, mobile applications have been regarded as substitutions of websites in many studies [127]. Large technology companies such as Facebook and Amazon have been focusing on user growth on mobile devices [78].

However, interventions may result in unintended outcomes [57, 55, 123]. Specifically, while some interventions may be highly effective at achieving the measured goal of a behavioral change system, they may reduce desired outcomes elsewhere [55]. In one health-related intervention, while the physical activity of participants increased, calorie intake also increased, working against the goal of promoting a healthy lifestyle [21]. Similarly, using peer pressure to build confidence for students at school would, in turn, lower their self-esteem which actually was opposite to the goal of augmenting confidence [123].

In our system, the time spent on unproductive activities might be decreased in one application yet increased in others. These prompt our hypotheses:

Hypothesis 7 (H7)**.** *Within a single device, productivity interventions will cause the time spent on targeted sites and apps to be redistributed to other sites and apps.*

Hypothesis 8 (H8)**.** *Between computers and mobile devices, productivity interventions will cause the time spent on one device to be redistributed to other devices.*

5.3 Experiment Platform: HabitLab

We conducted the studies in this chapter using the browser and android versions of HabitLab. At the time the studies presented in this chapter were conducted, the Chrome version had over 8000 daily active users, and the Android version had over 500 daily active users. The list of interventions that were included in HabitLab at the time of this study is included at the end of this chapter.

5.4 Study: Redistribution of Time Within and Across Devices

In this study we aim to analyze whether productivity interventions are reducing or redistributing time. We pursue this through an experiment and three sets of analyses: *(1) Within-device redistribution of time, in the browser*. For example, this would be the effects on time spent on non-Facebook websites, due to interventions that run when visiting the Facebook website. *(2) Within-device redistribution of time, on mobile devices.* For example, this would be the effects of time spent on non-Facebook applications, due to interventions that run when using the Facebook app. *(3) Cross-device redistribution of time.* For example, this would be the effects of time spent on Facebook on the phone, due to interventions that run when visiting the Facebook website.

5.4.1 Participants

Participants in this study consisted of new HabitLab users who installed either the HabitLab Chrome extension or Android app over a period of 132 days (approximately 19 weeks) in July through December 2018. 3747 users installed the HabitLab Chrome version over the course of our experiment and consented to our research protocol. 1483 users did so for the Android version. 298 installed both and signed in with their Google accounts, allowing us to analyze their usage across devices. We discarded participants who were not new users of HabitLab, since some users were re-installs or new devices for existing users. We also discarded participants who did not complete the onboarding process, or who uninstalled the system before they saw their first intervention. This left us with 1790 participants for Chrome, 782 participants for Android, and 82 participants for whom we could analyze usage across both. A summary of our dataset is shown in Table 5.1.

5.4.2 Method

In order to observe time redistribution effects between a focal goal and other goals due to interventions, we would ideally randomly turn interventions on and off for goals, then observe the effects on other goals.

Table 5.1: Data Summary. Note that the duration of 132 days are users who kept it installed the longest, but as users can freely install/uninstall we do not have 132 days of data on all users.

	Browser	**Android**	**Synced**
Time Duration	132 days	132 days	132 days
No. of Users	1790	782	82
No. of Sessions	4.8 million	11.3 million	3.8 million

However, because HabitLab informs users that it will show interventions on goals that they select, there would be negative consequences (e.g., user confusion and dissatisfaction) if interventions for an application disappeared entirely for a week. Therefore, we opt to vary frequency rather than entirely turn off interventions for a goal each week.

So, for each goal on each device, we randomize frequency of interventions each week. On weeks where a goal is set as frequent, an intervention is shown on every visit to the app or site. On weeks where a goal is set as infrequent, an intervention is shown with probability 0.2 on every visit to the app or site. We choose this methodology of varying frequency to approximate the effects of turning interventions entirely on or off.

We analyze the effects interventions have on overall time spent on goals in the browser and mobile environments. We do so with a linear mixed model, which models the relationship between a dependent variable of time spent that day on a goal, an independent variable of goal frequency (frequent or infrequent), and categorical variables for the user and the goal site or app (e.g., Facebook, YouTube, Reddit) as random effects. We run the model separately on both the data from the browser and mobile versions. Our results here can also be replicated with a simpler model of an independent sample t-test modeling the effects of frequency on time spent.

5.4.3 Intensity

Frequency measures how much a user is being nudged in a single goal, but our experiment also needs to measure how much a user is being nudged overall, across all goals on the platform. This allows us to, for example, measure whether mobile device usage increases when browser interventions are overall more frequent, or whether time spent on non-goal sites increases when interventions are more frequent on goal sites. So, we define a measure of *intensity*: the percentage of sessions on any goal that triggered an intervention. For example, if the goal apps are Facebook and YouTube, the user visited Facebook 10 times and saw interventions 2 times, and visited YouTube 3 times and saw interventions 3 times, then the intensity is $\frac{5}{13} = .38$. Intensity will naturally vary over time as goals are re-randomized into *frequent* and *infrequent* conditions, with more frequent goals increasing intensity and more infrequent goals decreasing intensity. This randomization occurs for all goals simultaneously, once a week. We chose this intensity metric for our analysis, as opposed to alternatives such as raw number of times interventions were seen, because: 1) it is independent of the dependent variable, total time spent; 2) it is independent of the number of times the user visits a site/app; 3) it is guaranteed to be between 0 to 1, which is useful for interpretation; and 4) it can be used for both within-device and cross-device analysis.

For each goal, we also define a measure of *intensity of other goals.* This is the intensity measure excluding the current goal. We will use it for analyzing redistribution of time within device: when intensity of other goals varies, what is the effect on time spent on a target goal?

5.4.4 Time Redistribution

Within Device

We analyze the effects of interventions on time redistribution within device. We define *time redistribution within device* as an increase in time spent on the goal on the device, as a result of a change in intensity of other goals. For example, an increase in time spent on YouTube as a result of turning Facebook interventions on would be an example of time redistribution from Facebook to YouTube.

We do so with a linear mixed model, which models the relationship between a dependent variable of time spent that day on all goals, an independent variable of intensity of goals, as well as the user as a random effect. We run the model separately on both the data from the browser and mobile versions. Because our time data is log-normally distributed, we fit our linear mixed models to log time.

Across Device

We analogously define time redistribution between devices as an increase in time spent on the other device, as a result of interventions increasing in frequency in the other device. For example, an increase in time spent on Facebook on the browser, as a result of increasing the frequency of interventions on mobile would be an example of time being redistributed from mobile to browser.

We do so with a linear mixed model, which models the relationship between a dependent variable of time spent that day on all goals on one device, an independent variable of intensity of goals on the other device, and the user as a random effect. We run the model separately on data in both directions: one analyzing the effects of browser intensity on time spent on mobile, and another analyzing the effects of mobile intensity on time spent on the browser. We again log transform our time data for analysis.

5.5 Results

First, we establish that our interventions are effective – that is, increasing the frequency of intervention on a goal app reduces time on that app. Next, we confirm that increasing intensity on a device reduces time on goal apps on that device. Then, we analyze redistribution effects within device – that is, whether increasing intensity effects time on non-goal apps. We also analyze redistribution effects across devices – that is, whether increasing intensity on one device effects time on goal apps on the other device. Finally, we build intuition for the underlying mechanisms by exploring what happens after users visit goal applications.

Table 5.2: Browser: Frequent interventions for a goal site cause a reduction of time spent on the site.

	Dependent variable:
	Log daily time on site
Frequent (1=true)	-0.085^{***}
	(0.010)
Baseline	5.904^{***}
	(0.224)
Observations	96,489
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5.3: Mobile: Frequent interventions for a goal app cause a reduction of time spent on the app.

	Dependent variable:
	Log daily time on app
Frequent (1=true)	-0.045^{***}
	(0.011)
Baseline	5.254^{***}
	(0.057)
Observations	96,147
Note:	*p<0.05; **p<0.01; ***p<0.001

5.5.1 Are interventions effective?

Browser

Yes. We look at the effect of frequency of interventions on time spent on a day on a site, controlling for the user and the goal. We find a significant reduction in time spent on day on an app, when interventions for that goal are frequently shown that day ($p < 0.001$), as shown in Table 5.2. Estimated log time on a goal when infrequent is 5.747 (313 seconds), while for frequent goals this is reduced to 5.665 (288 seconds). Hence, our methodology of increasing intervention frequency is effective at reducing time on sites.

Mobile

Yes. We look at the effect of frequency of interventions on time spent on a day on an app, controlling for the user and the goal. We find a significant reduction in time spent on day on an app, when interventions for that goal are frequently shown that day ($p < 0.001$), as shown in Table 5.3. Estimated log time on a goal when infrequent is 5.928 (375 seconds), while for frequent goals this is reduced to 5.462 (235 seconds). Hence, our methodology of increasing intervention frequency is effective at reducing time on apps.

Table 5.4: Browser: Increasing intensity results in a reduction of time spent each day on all goal domains

	Dependent variable:
	Log daily time spent on all goal sites
Browser Intensity	-0.187^{***}
	(0.016)
Baseline	6.929^{***}
	(0.033)
Observations	57,204
Note:	*p<0.1; **p<0.05; ***p<0.01

5.5.2 Is time spent on goals reduced when there is higher intensity?

In the previous analysis we have shown that increasing frequency of interventions on a single goal allows us to observe reductions in time spent on that goal, on both the browser and mobile platforms. In this section we will show that increasing intensity also allows us to observe reductions in total time spent on all goal apps, on both platforms. This allows us to confirm the validity of our intensity metric, as well as allow us to analyze the aggregate usage of all goal apps on each device. This will be necessary for our later analyses of redistribution effects within device as well as between devices.

Browser

Yes. We look at the effect of intensity on total time spent on goal sites each day, controlling for the user. We find a significant reduction in total time spent on goal sites when intensity is higher that day ($p < 0.001$), as shown in Table 5.4. Estimated log total time on goal sites with low intensity is 6.885 (978 seconds), while with high intensity this is reduced to 6.758 (861 seconds). Hence, when interventions are more frequent in aggregate on the browser (which intensity captures), overall time on goal sites is reduced.

Mobile

Yes. Like the browser, we look at the relationship between increasing intensity on one's mobile phone and the total time spent that day on one's goal applications. We find a significant decrease ($p < .05$) in goal time spent, as shown in Table 5.5. Estimated log total time on goal apps with low intensity is 8.146 (3450 seconds), while with high intensity this is reduced to 8.031 (3075 seconds). Hence, when interventions are more frequent in aggregate on mobile (which intensity captures), overall time on goal apps is reduced.

Table 5.5: Mobile: Increasing intensity results in a reduction of time spent each day on all goal apps

	Dependent variable:
	Log daily time spent on all goal apps
Mobile Intensity	-0.049^{*}
	(0.025)
Baseline	8.300^{***}
	(0.042)
Observations	22,970
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 5.6: Browser: Increasing intensity results in a reduction of time spent each day on non-goal sites

	Dependent variable:
	Log daily time spent on all non-goal sites
Browser Intensity	-0.169^{***}
	(0.016)
Baseline	8.207^{***}
	(0.028)
Observations	57,204
Note:	*p<0.1; **p<0.05; ***p<0.01

5.5.3 What is the effect of increasing intensity on other, non-goal apps and sites?

Browser

Reduction. We look at the effect of intensity on total time spent on non-goal sites each day, controlling for the user. We find a significant reduction in total time spent on non-goal sites when intensity is higher that day ($p < 0.000005$), as shown in Table 5.6. Estimated log total time on non-goal sites when intensity=0 is 8.207 (3667 seconds), while when intensity=1 this is reduced to 8.038 (3096 seconds). This is the effect predicted by our global reduction hypothesis.

Mobile

No effect (isolation). We do not observe a significant effect of Android intensity on time outside of goals, as shown in 5.7. This suggests that reducing time within Android is an "isolated" behavior. Note there is an insignificant trend towards increasing time on non-goal sites with increasing intensity ($p = 0.07$).

Table 5.7: Mobile: Increasing intensity has no significant effect of time spent on non-goal apps.

	Dependent variable:
	Log daily time spent on non-goal apps
Mobile Intensity	0.035
	(0.020)
Baseline	9.277^{***}
	(0.044)
Observations	22,970
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 5.8: Mobile: Varying intervention intensity has no effect on total time spent on browser goal sites

	Dependent variable:
	Log daily time spent on browser goals
Mobile Intensity	0.045
	(0.218)
Baseline	6.736^{***}
	(0.251)
Observations	1,312
Note:	*p<0.05; **p<0.01; ***p<0.001

5.5.4 Is time redistributed between devices?

Mobile to Browser

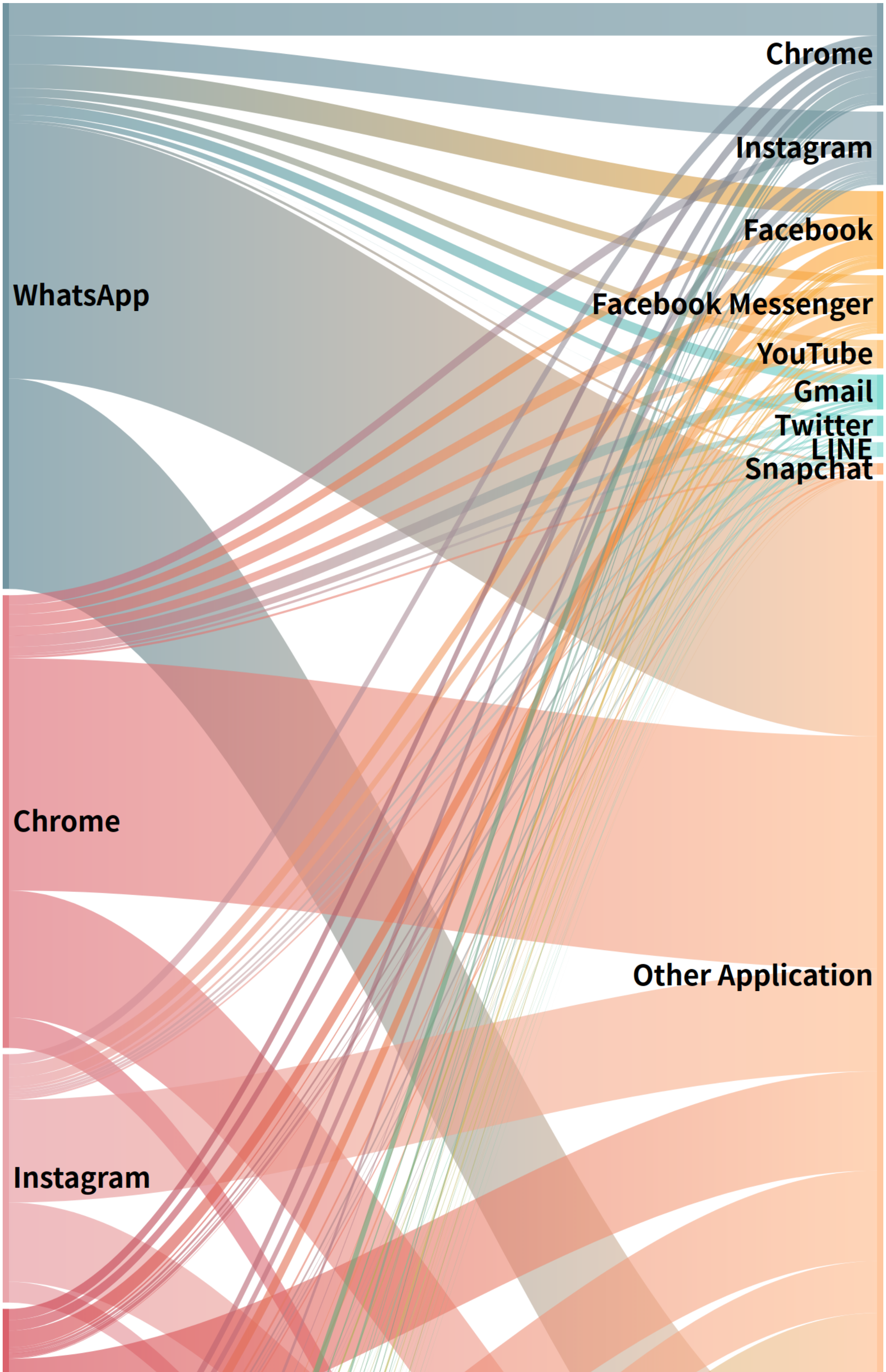
No effect (isolation). We look at the effect of mobile intervention intensity, on total time spent on browser. We find no significant effect (p¿.5), as shown in Table 5.8.

Browser to Mobile

No effect (isolation). We look at the effect of browser intervention intensity, on total time spent on mobile. We find no significant effect (p¿.5), as shown in Table 5.9.

5.5.5 Destination tracking

Finally, to build intuition as to the mechanism by which the above effects are happening, we analyzed what happens after users their goal applications. We visualized the flow of sessions from the 10 most widely chosen goal apps and sites in our dataset as Sankey diagrams (Figures 5.3 and 5.4). On mobile, a majority of sessions end up going to another application, followed by turning off the phone, as shown in Figure 5.3. On browsers, the majority of sessions went to other sites, as shown in Figure 5.4. We can also observe


WhatsApp
Chrome
Instagram
Chrome
Instagram
Facebook
Facebook Messenger
YouTube
Gmail
Twitter
LINE
Snapchat
Other Application

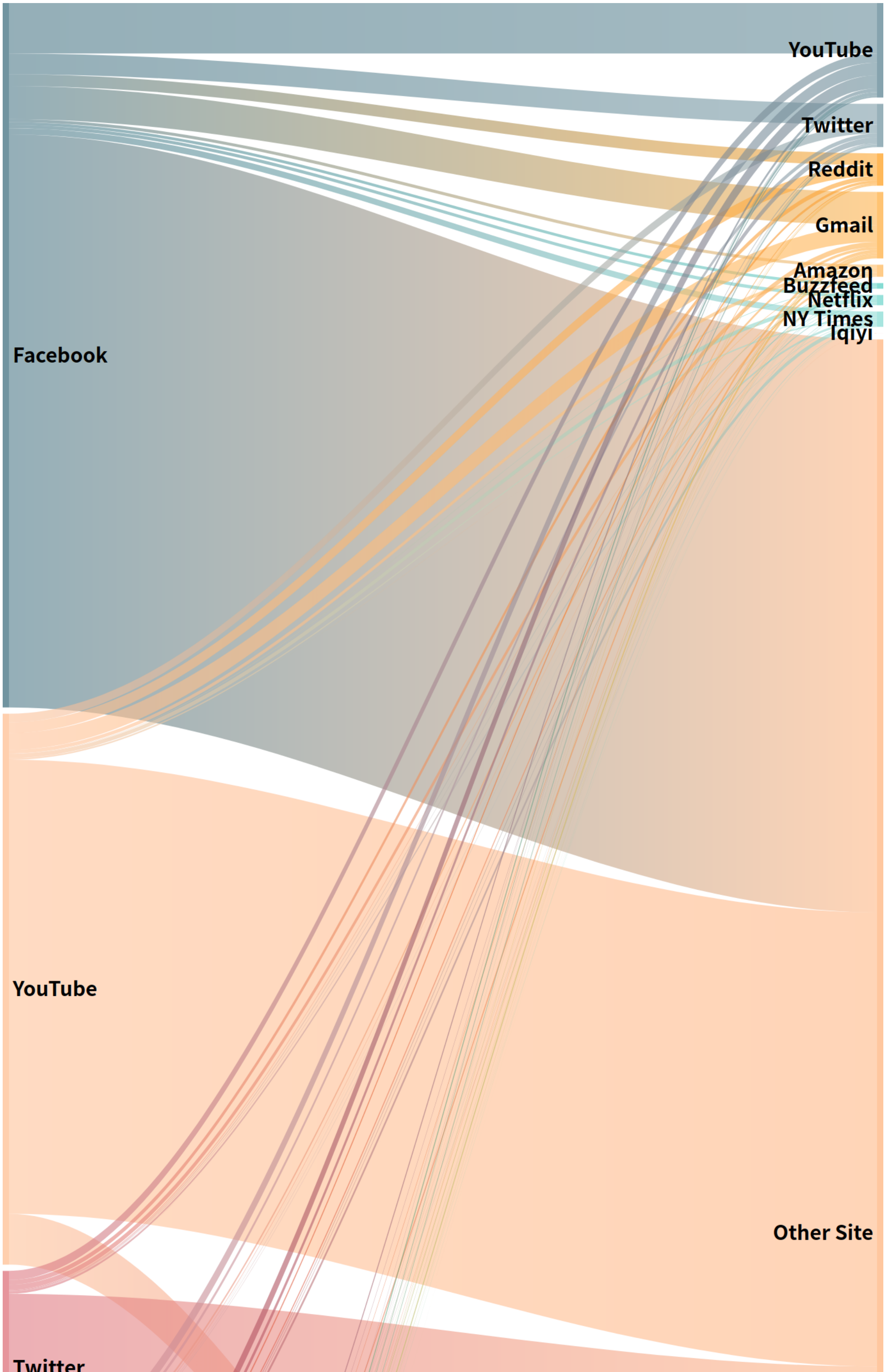

Facebook
YouTube
Twitter
YouTube
Twitter
Reddit
Gmail
Amazon
Buzzfeed
Netflix
NY Times
Iqiyi
Other Site

Table 5.9: Browser: Varying intervention intensity has no effect on total time spent on mobile goal apps

	Dependent variable:
	Log daily time spent on mobile goals
Browser Intensity	0.064
	(0.068)
Constant	8.219^{***}
	(0.128)
Observations	1,312
Note:	*p<0.05; **p<0.01; ***p<0.001

differences in goals users choose on mobile as opposed to desktop – on mobile, the most popular apps tend to be messaging apps, whereas on the browser they tend to be content aggregators.

5.6 Limitations

Our methodology varied frequency of interventions, instead of comparing having interventions completely on vs completely off. This approach reduces the size of effects we can observe compared to having interventions completely on or completely off. Our approach is also sensitive to variance in the effectiveness levels of the interventions. Some interventions may be more aggressive than others and change users' behavior more drastically even with low frequency. This difference may alter time re-distributions due to varied frequency.

We did not measure time spent on platforms that HabitLab does not support. For instance, HabitLab users may use Facebook on tablet devices, watch TV or engage in other activities that are considered unproductive aside from browsing on a desktop or on an Android phone. These behaviors may potentially change how time redistributed, but we are unable to track it.

Additionally, our study explores time redistribution in the context of productivity. It is possible that this context may not generalize to other behavior change regimes.

5.7 Discussion

We found that productivity interventions on the browser also reduced time on sites other than the targeted sites, but there was no such effect on mobile or cross-device.

We believe the reason we observed reduction in time on non-goal sites on the browser is several of the most popular goal sites — such as Facebook, Reddit, Twitter — are filled with hyperlinks to other sites, and hence drive traffic to them. For example, if an intervention makes a user spend less on their Facebook feed, they are going to be less likely to stumble upon a New York Times article, hence the Facebook-reducing intervention may also reduce time on New York Times. Part of this may be a difference in how mobile applications work, compared to websites. Several mobile applications embed a web browser so that even if the user clicks a link, it will open within the same app. For example, Facebook is one such app, so if the user clicks on a New York Times link within the Facebook app, it is opened within the Facebook app's built-in browser, so the time they spend reading that article will still be counted towards Facebook app usage.

One possible reason for differences between mobile and web is that the apps users choose to reduce time on in each two platform differ (e.g., messaging apps on Android vs. link aggregators on Chrome). There also exist differences in typical interaction styles (short, notification-driven sessions on Android [98], vs. longer sessions resulting from self-interruption on Chrome). 85% of the apps that Android users frequently chose to reduce time on are for messaging (WhatsApp, Instagram, Facebook Messenger, Twitter, LINE, Snapchat), where a characteristic interaction is receiving a message, unlocking the phone to read it and reply, then turning off the screen (as shown in Figure 5.3). Thus, users would not be drawn to other apps during this interaction. In contrast, with the Chrome version, the most selected sites are Facebook, YouTube, Reddit, and Twitter, 75% of which are aggregators of links to other sites. The number of daily sessions per app is also greater on Android, though sessions are longer on average on Chrome, and stopping using the browser after a session ends occurs less on Chrome. Thus, the browser-based interactions users were using HabitLab to reduce are not short messaging-driven spurts that end with turning off the screen as on mobile, but rather long sessions of surfing through link aggregators ending with going to another site. So, a proposed mechanism: interventions short-circuit browsing long browser-based sessions, but mobile sessions are already short.

This work brings about implications for designing interventions. Namely, we should consider not only the immediate interaction and its immediately measurable effects, but its longer-term effects in the context of the broader workflow. For example, consider 2 interventions for Facebook: 1) asks users to return to the home screen, vs 2) asks users to turn off the screen. Assuming similar rates of compliance, we would expect that measuring the effects on time spent on Facebook in isolation will show no difference between them. However, if we consider that going to the home screen can lead to users opening other apps, we might predict that a holistic measurement that includes effects on other apps as well will prefer 2) over 1). Or if designing interventions to reduce snacking, should we: a) ask participants to not eat anything until their next meal, or b) give them gum instead? While calorie intake from the immediate interaction would favor a), b) may prevent future snacking down the line. That said, in many cases, interventions are indeed isolated in their effects, and can even have beneficial effects elsewhere.

5.8 Conclusion

In this paper we have compared three hypotheses for how productivity interventions influence time spent on sites, apps, and devices other than the ones they are targeting. Productivity interventions may have no effect on other goals (*isolated effects*), they may cause time to be redistributed to other unproductive goals (*redistribution*), or they may cause a reduction in time spent on other unproductive goals (*reduction*).

We adjudicated between these hypotheses by varying the frequency of productivity interventions on goals that users set in the HabitLab browser extension and mobile app. When interventions were more frequent, users spent less time on their goal sites and apps, showing that the productivity interventions were effective. We also defined a metric of intensity that captures frequency of interventions within device, and investigated the effects of varying intensity of interventions for other apps/sites, on time spent on an app/site. The result differed by device: on the browser we observed a global reduction effect, with time on non-goal sites decreasing with increasing intensity of interventions. However, on mobile we observed no effect. We believe these differences are caused by differing usage patterns and platform differences: websites drive traffic to other websites via hyperlinks, but mobile apps try to keep users remaining on the app.

We have shown that while productivity interventions can sometimes have effects on usage of other, non-targeted sites and apps, they are often isolated in their effects. Hence, when designing for behavior change, while we should be careful about our measurements and the possibility of unintended side effects, in the context of productivity interventions it appears that targeting individual productivity goals does not cause substantial negative second-order effects.

5.9 Appendix: List of Browser Interventions used in this study

The following is the list of interventions used for this study, showing the intervention name and description as seen by the end user.

Generic interventions that can be used on all sites:

- Minute Watch: Notifies you of time spent every minute
- Supervisor: Shows time spent on site at the top of screen
- Scroll Freezer: Freezes scrolling after a certain amount of scrolls
- Stat Whiz: Show time spent and visit count each visit
- GateKeeper: Makes you wait a few seconds before visiting
- 1Min Assassin: Closes tab after 60 seconds
- Bouncer: Asks how long you want to spend on site this visit

Facebook-specific interventions:

- Time Injector: Injects timer into the Facebook feed
- Feed Eater: Removes the Facebook news feed
- TimeKeeper: Notifies you of time spent in the corner of your desktop
- No Comment: Removes Facebook comments
- Clickbait Mosaic: Removes clickbait from the news feed

Youtube-specific interventions:

- Sidekicker: Remove sidebar links
- Think Twice: Prompt the user before watching a video
- No Comment: Removes comment section

Netflix-specific interventions:

- Fun Facts: Gives you a fact and links an article on the effect of TV
- Alarm Clock: Asks the user to set an alarm before watching a show
- Stop Autoplay: Stops the site from automatically playing the next video

Reddit-specific interventions:

- Comment Remover: Removes Reddit comments
- Mission Objective: Asks what you aim to do this visit and puts a reminder up

Youku-specific interventions

- Think Twice: Prompt the user before watching a video
- Sidekicker: Remove sidebar links

iQiyi-specific interventions

- Think Twice: Prompt the user before watching a video
- Sidekicker: Remove sidebar links

Twitter-specific interventions:

- Feed Eater: Removes the Twitter news feed

Amazon-specific interventions:

- No Recs: Hides recommendations

Gmail-specific interventions

- Speedbump: Delays the arrival of new emails

5.10 Appendix: List of Mobile Interventions used in this study

All mobile interventions are generic, that is they can be used on any app.

- At it Again: Sends a pop up with your app visit count.
- Progress Report: Sends a pop up with today's total usage for a certain app
- Red Alert!: Sends a notification with today's total usage for a certain app
- Repeat Offender: Sends a notification with your app visit count
- All in All: Pops a dialog with the day's total time on the current app
- Back To Target: Suggests you to visit a target app
- Counting on You: Puts a timer on screen in watchlisted apps
- Man Overboard! Shows a dialog with your app visit count
- No Peeking!: Asks for confirmation before opening watchlisted apps
- Wait Up! Pause for 10 seconds before entering an app
- Your Better Half: Sends a pop up to go to a target app
- Look on the Bright Side: Dim the screen a little at a time
- Take Your Pick: Select how long you want to spend on an app
- The Final Countdown: On screen timer that closes the app when time runs out

The following interventions apply across the device as a whole, not individual applications.

- How Time Flies!: Sends a pop up message with current app visit length
- Knock Knock: Sends a pop up with your glance count for the day
- Long Time No See: Sends pop up with your phone usage for the day
- Call it a Day: Sends notification with phone usage for the day
- Easy on the Eyes: Sends notification with glance count for the day
- Hello, Old Friend: Sends notification with unlock count for the day
- The Clock is Ticking: Sends a notification with the current app visit duration
- En Garde: Pops a dialog with the day's total unlock count
- Hold the Phone: Show dialog with phone usage for the day
- Long Story Short: Pops a dialog with the visit time for the current app
- Quote reminder: Show quote upon opening app
- Time Reminder: Show dialog with phone usage for the day
- Take Your Pick: Select how long you want to spend on an app

Chapter 6

Discussion and Conclusion

In this thesis we have proposed a paradigm of in-the-wild experimentation to gain insights about behavior change, and have created a platform, HabitLab, to realize this vision in the context of helping users reduce their time online and on their phones. We also conducted a set of studies on HabitLab which illustrate that we can make novel findings with the system.

Our design principles with HabitLab focused on maximizing user retention rates, which we believe helped us with growth. We did so by giving users control of their interventions, providing visually appealing and unobtrusive interventions, and avoiding intrusive surveys and experience sampling as much as possible.

A major advantage we had with the domain we chose – online behavior change – is that we could push our interventions automatically to users on every visit, without requiring any interaction on their part. This may be considerably more difficult to realize in other behavior change domains, where it may be more difficult to sense when the targeted activity is taking place. For instance, dieting apps unable to detect when the user is about to eat may require the user to explicitly open the app to indicate when and what they are eating, so different design strategies – such as unprompted push notifications that some might consider to be intrusive – may be needed for other behavior change domains.

The first set of studies we ran with HabitLab investigated whether interventions decline in effectiveness over time. We found that interventions decline in effectiveness if the same intervention is repeatedly shown, and that a strategy of rotating between different interventions can help improve the effectiveness. While this comes at the cost of increased attrition, most likely due to users having incorrect mental models, we can reduce this attrition via a simple design shown when a new intervention is introduced.

We believe that novelty is an underlying mechanism for the improvement in effectiveness we observed when interventions are rotated. This leads us to speculate: would it be a effective and practical strategy to scale up the number of interventions, so that we are rotating between interventions from a huge pool of hundreds of interventions? Or do the improvements in effectiveness that we can expect from rotating increasing numbers of interventions have limits? We speculate that increasing the number of interventions in rotation will have high initial benefits for the first few additional interventions, but will have declining benefits as more interventions are added, as the probability of repeatedly seeing a recently-seen intervention grows increasingly small.

The second set of studies investigated whether interventions that help save time on one site, app, or device influence time spent elsewhere. We found that on the browser, reducing time on one site has a beneficial side effect of reducing time elsewhere. We believe this is due to reducing time on aggregator sites that drive traffic to other sites. On phones, however, we did not observe any side effects of reducing time on one app on other apps. We also did not observe any cross-device effects.

The findings of this study have a positive tone – we did not observe negative side effects of productivity interventions, which would have been predicted if users were using their devices to replenish willpower when exhausted. Perhaps one speculative explanation is that in the context of device usage, diminished willpower results in the user opening a site or visiting an app, but actually spending time on the site or app does not replenish this willpower – only the initial act of opening the site or app does. This may have interesting implications if it is true in other domains as well – for instance, if the user is on a diet and has a craving for doughnuts, would an intervention preventing them from eating a doughnut also suppress cravings for other fattening foods as well? If they give in to the craving, would stopping them after the first bite leave their craving equally satisfied as if we let them eat the whole doughnut?

Given we conducted these studies in the context of reducing time spent online and on phones, results may not necessarily generalize to other behavior change domains. That said, with our increasing ability to sense our environment via sensors in our phones, smartwatches, and IoT devices, many of the paradigms we used in HabitLab can be applied to other domains as well. For instance, in the fitness domain, if we can sense users' physical activity levels via a smartwatch, we can experiment with various interventions that prompt users to exercise by playing audio messages or sending notifications. This hypothetical in-the-wild experimentation platform for fitness could potentially work analogously to HabitLab, running studies to find intervention strategies that work well to increase physical activity levels. The increasing ubiquity of sensors in the physical world make this paradigm of in-the-wild behavior change increasingly realistic and possible in domains outside online behavior change.

There is a large opportunity for behavior change research through big data and crowdsourcing that has been under-explored due to the paucity of large-scale deployments of research systems. Could we predict which interventions will work well for a new user, before they even start using the system? Could we automatically deploy and test modified versions of interventions, to hill-climb our way to more effective ones? Could we enlist an engaged user community to come up with, generate, and test new interventions for the long-tail of behavior change goals that designers had never even thought of? These can be realized with machine learning and crowdsourcing techniques, but there have not been appropriate communities for an in-the-wild deployment. We hope HabitLab will provide such a platform to realize this vision of community-driven behavior change research in the wild.

Bibliography

[1] Yielding to (cyber)-temptation: Exploring the buffering role of self-control in the relationship between organizational justice and cyberloafing behavior in the workplace. *Journal of Research in Personality*, 45(2):247 – 251, 2011.

[2] Charles Abraham and Susan Michie. A taxonomy of behavior change techniques used in interventions. *Health psychology*, 27(3):379, 2008.

[3] Phil Adams, Mashfiqui Rabbi, Tauhidur Rahman, Mark Matthews, Amy Voida, Geri Gay, Tanzeem Choudhury, and Stephen Voida. Towards personal stress informatics: Comparing minimally invasive techniques for measuring daily stress in the wild. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pages 72–79. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.

[4] Eytan Adar, Desney S Tan, and Jaime Teevan. Benevolent deception in human computer interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1863–1872. ACM, 2013.

[5] Elena Agapie, Daniel Avrahami, and Jennifer Marlow. Staying the course: System-driven lapse management for supporting behavior change. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1072–1083. ACM, 2016.

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

[7] Icek Ajzen and Martin Fishbein. Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological bulletin*, 84(5):888, 1977.

[8] AR Allgeier, Donn Byrne, Barbara Brooks, and Diane Revnes. The waffle phenomenon: Negative evaluations of those who shift attitudinally1. *Journal of Applied Social Psychology*, 9(2):170–182, 1979.

[9] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Steering user behavior with badges. In *Proceedings of the 22nd international conference on World Wide Web*, pages 95–106. ACM, 2013.

[10] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide web*, pages 687–698. ACM, 2014.

[11] Cecilie Schou Andreassen, Torbjørn Torsheim, Geir Scott Brunborg, and Ståle Pallesen. Development of a facebook addiction scale. *Psychological reports*, 110(2):501–517, 2012.

[12] Lorraine M. Angelino, Frankie Keels Williams, and Deborah Natvig. Strategies to engage online students and reduce attrition rates. *Journal of Educators Online*, 4:14, 07 2007.

[13] Steven H. Appelbaum, Adam Marchionni, and Arturo Fernandez. The multi-tasking paradox: perceptions, problems and strategies. *Management Decision*, 46(9):1313–1325, 2008.

[14] Azy Barak, Liat Hen, Meyran Boniel-Nissim, and Na'ama Shapira. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *Journal of Technology in Human Services*, 26(2-4):109–160, 2008.

[15] Alan Baron, Michael Perone, and Mark Galizio. Analyzing the reinforcement process at the human level: can application and behavioristic interpretation replace laboratory research? *The Behavior Analyst*, 14(2):95, 1991.

[16] Roy E Baumeister, Ellen Bratslavsky, Mark Muraven, and Dianne M Tice. Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74(5):1252–1265, 1998.

[17] Chris Beer. Multi-device watching among tv viewers, 2017.

[18] Jan Panero Benway. Banner blindness: The irony of attention grabbing on the world wide web. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 42, pages 463–467. SAGE Publications Sage CA: Los Angeles, CA, 1998.

[19] Michel Bernier and Jacqueline Avard. Self-efficacy, outcome, and attrition in a weight-reduction program. *Cognitive Therapy and Research*, 10(3):319–338, Jun 1986.

[20] Ashly D. Black, Josip Car, Claudia Pagliari, Chantelle Anandan, Kathrin Cresswell, Tomislav Bokun, Brian McKinstry, Rob Procter, Azeem Majeed, and Aziz Sheikh. The impact of ehealth on the quality and safety of health care: A systematic overview. *PLOS Medicine*, 8(1):1–16, 01 2011.

[21] S N Blair, D R Jacobs, and K E Powell. Relationships between exercise or physical activity and other health behaviors. 100:172 – 180, 03 1985.

[22] L. Mark Carrier, Nancy A. Cheever, Larry D. Rosen, Sandra Benitez, and Jennifer Chang. Multitasking across generations: Multitasking choices and difficulty ratings in three generations of americans. *Computers in Human Behavior*, 25(2):483 – 489, 2009. Including the Special Issue: State of the Art Research into Cognitive Load Theory.

[23] Stevie Chancellor, Andrea Hu, and Munmun De Choudhury. Norms matter: Contrasting social support around behavior change in online weight loss communities.

[24] Eun Kyoung Choe, Sunny Consolvo, Nathaniel F Watson, and Julie A Kientz. Opportunities for computing technologies to support healthy sleep behaviors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3053–3062. ACM, 2011.

[25] Yu-Kai Chou. *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. 2015.

[26] Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A. Munson. When personal tracking becomes social: Examining the use of instagram for healthy eating. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 1674–1687, New York, NY, USA, 2017. ACM.

[27] Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine Zia, James Fogarty, Julie A. Kientz, and Sean A. Munson. Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 770–786, New York, NY, USA, 2016. ACM.

[28] Robert B Cialdini and Nathalie Garde. *Influence*, volume 3. A. Michel, 1987.

[29] Emily IM Collins, Anna L Cox, Jon Bird, and Daniel Harrison. Social networking use and rescuetime: the issue of engagement. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 687–690. ACM, 2014.

[30] The Nielsen Company. Television, internet and mobile usage in the us. three screen report, 2009.

[31] Sunny Consolvo, David W McDonald, and James A Landay. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 405–414. ACM, 2009.

[32] Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. Activity sensing in the wild: A field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1797–1806, New York, NY, USA, 2008. ACM.

[33] Alexander P. Cotter, Nefertiti Durant, April A. Agne, and Andrea L. Cherrington. Internet interventions to support lifestyle modification for diabetes management: A systematic review of the evidence. *Journal of Diabetes and its Complications*, 28(2):243 – 251, 2014.

[34] Brian Cugelman. Gamification: what it is and why it matters to digital health behavior change developers. *JMIR Serious Games*, 1(1):e3, 2013.

[35] Pim Cuijpers, Annemieke van Straten, and Gerhard Andersson. Internet-administered cognitive behavior therapy for health problems: a systematic review. *Journal of Behavioral Medicine*, 31(2):169–177, Apr 2008.

[36] Laura Dabbish, Gloria Mark, and Víctor M González. Why do i keep interrupting myself?: environment, habit and self-interruption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3127–3130. ACM, 2011.

[37] Kathryn P Davison, James W Pennebaker, and Sally S Dickerson. Who talks? the social psychology of illness support groups. *American Psychologist*, 55(2):205, 2000.

[38] Walter Dempsey, Peng Liao, Pedja Klasnja, Inbal Nahum-Shani, and Susan A. Murphy. Randomised trials for the fitbit generation. *Significance*, 12(6):20–23.

[39] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15. ACM, 2011.

[40] P Dolan, M Hallsworth, D Halpern, D King, and I Vlaev. Mindspace: Influencing behaviour through public policy. 2010. *Cabinet Office and Institute for Government, London, UK.*

[41] Charles Duhigg. *The power of habit: Why we do what we do in life and business*, volume 34. Random House, 2012.

[42] Daniel A Epstein, Felicia Cordeiro, James Fogarty, Gary Hsieh, and Sean A Munson. Crumbs: Lightweight daily food challenges to promote engagement and mindfulness. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 2016.

[43] Daniel A Epstein, Bradley H Jacobson, Elizabeth Bales, David W McDonald, and Sean A Munson. From nobody cares to way to go!: A design framework for social sharing in personal informatics. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1622–1636. ACM, 2015.

[44] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 731–742. ACM, 2015.

[45] Nir Eyal. *Hooked: How to build habit-forming products.* Penguin Canada, 2014.

[46] Gunther Eysenbach. The law of attrition. *Journal of medical Internet research*, 7(1), 2005.

[47] Gunther Eysenbach, John Powell, Marina Englesakis, Carlos Rizo, and Anita Stern. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *Bmj*, 328(7449):1166, 2004.

[48] Leon Festinger. A theory of social comparison processes. *Human relations*, 7(2):117–140, 1954.

[49] Brianna S. Fjeldsoe, Alison L. Marshall, and Yvette D. Miller. Behavior change interventions delivered by mobile telephone short-message service. *American Journal of Preventive Medicine*, 36(2):165 – 173, 2009.

[50] Focus Booster. Focus booster.

[51] Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):5, 2002.

[52] David Freedman. What's next: Taskus interruptus, 2007.

[53] Jon Froehlich, Tawanna Dillahunt, Predrag Klasnja, Jennifer Mankoff, Sunny Consolvo, Beverly Harrison, and James A Landay. Ubigreen: investigating a mobile tool for tracking and supporting green transportation habits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1043–1052. ACM, 2009.

[54] J. Hamari, J. Koivisto, and H. Sarsa. Does gamification work? – a literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on System Sciences*, pages 3025–3034, Jan 2014.

[55] Juho Hamari, Jonna Koivisto, and Tuomas Pakkanen. Do persuasive technologies persuade? - a review of empirical studies. In Anna Spagnolli, Luca Chittaro, and Luciano Gamberini, editors, *Persuasive Technology*, pages 118–136, Cham, 2014. Springer International Publishing.

[56] W Hardeman, S Griffin, M Johnston, AL Kinmonth, and NJ Wareham. Interventions to prevent weight gain: a systematic review of psychological models and behaviour change methods. *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity*, 24(2):131–143, 2000.

[57] Bernard J Healey and Robert S Zimmerman. *The new world of health promotion : new program development, implementation, and evaluation.* Jones and Bartlett Publishers, 2010.

[58] A. Hiniker, S. Heung, S. Hong, and J.A. Kientz. Coco's videos: An empirical investigation of video-player design features and children's media use. In *Proceedings of the 36th annual ACM conference on human factors in computing systemss*, CHI '18, New York, NY, USA, 2018. ACM.

[59] Alexis Hiniker, Sungsoo (Ray) Hong, Tadayoshi Kohno, and Julie A. Kientz. Mytime: Designing and evaluating an intervention for smartphone non-use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4746–4757, New York, NY, USA, 2016. ACM.

[60] Yan Hong, Ninfa C Pena-Purcell, and Marcia G Ory. Outcomes of online support and resources for cancer survivors: a systematic literature review. *Patient education and counseling*, 86(3):288–296, 2012.

[61] Jina Huh, Leslie S Liu, Tina Neogi, Kori Inkpen, and Wanda Pratt. Health vlogs as social support for chronic illness management. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(4):23, 2014.

[62] Kai Huotari and Juho Hamari. Defining gamification: a service marketing perspective. In *Proceeding of the 16th International Academic MindTrek Conference*, pages 17–22. ACM, 2012.

[63] Hasmida Jamaluddin, Zauwiyah Ahmad, Mazni Alias, and Maimun Simun. Personal internet use: The use of personal mobile devices at the workplace. *Procedia - Social and Behavioral Sciences*, 172:495 – 502, 2015. Contemporary Issues in Management and Social Science Research.

[64] Jing Jin and Laura A Dabbish. Self-interruption on the computer: a typology of discretionary task interleaving. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1799–1808. ACM, 2009.

[65] Pamela R. Johnson and Julie Indvik. The organizational benefits of reducing cyberslacking in the workplace. *Journal of Organizational Culture, Communications and Conflict*, 8(2), 2004.

[66] Berton H Kaplan, John C Cassel, and Susan Gore. Social support and health. *Medical care*, 15(5):47–58, 1977.

[67] Maurits Kaptein, Panos Markopoulos, Boris de Ruyter, and Emile Aarts. Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies*, 77:38–51, 2015.

[68] Ravi Karkar, Jessica Schroeder, Daniel A. Epstein, Laura R. Pina, Jeffrey Scofield, James Fogarty, Julie A. Kientz, Sean A. Munson, Roger Vilardaga, and Jasmine Zia. Tummytrials: A feasibility study of using self-experimentation to detect individualized food triggers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 6850–6863, New York, NY, USA, 2017. ACM.

[69] Matthew Kay, Eun Kyoung Choe, Jesse Shepherd, Benjamin Greenstein, Nathaniel Watson, Sunny Consolvo, and Julie A Kientz. Lullaby: a capture & access system for understanding the sleep environment. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 226–234. ACM, 2012.

[70] Vera Khovanskaya, Eric PS Baumer, Dan Cosley, Stephen Voida, and Geri Gay. Everybody knows what you're doing: a critical design approach to personal informatics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3403–3412. ACM, 2013.

[71] Junghyun Kim and Jong-Eun Roselyn Lee. The facebook paths to happiness: Effects of the number of facebook friends and self-presentation on subjective well-being. *CyberPsychology, behavior, and social networking*, 14(6):359–364, 2011.

[72] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, K Kim, and Jinwook Seo. Timeaware: Leveraging framing effects to enhance personal productivity. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 2016.

[73] Minsam Ko, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Daehee Shin, Koji Yatani, Junehwa Song, and Kyong-Mee Chung. 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*, CSCW '15, pages 1235–1245, New York, NY, USA, 2015. ACM.

[74] Reza Kormi nouri, Lars Goran Nilson, and Nobuo Ohta. The novelty effect: Support for the novelty encoding hypothesis. *Scandinavian Journal of Psychology*, 46(2):133–143.

[75] Geza Kovacs, Zhengxuan Wu, and Michael S Bernstein. Rotating online behavior change interventions increases effectiveness but also increases attrition. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), November 2018.

[76] Paul Krebs, James O Prochaska, and Joseph S Rossi. A meta-analysis of computer-tailored interventions for health behavior change. *Preventive medicine*, 51(3-4):214–221, 2010.

[77] Huitian Lei, Ambuj Tewari, and Susan A. Murphy. An actor-critic contextual bandit algorithm for personalized mobile health interventions. 06 2017.

[78] Amanda Lenhart, Kristen Purcell, Aaron Smith, and Kathryn Zickuhr. Social media and mobile internet use among teens and young adults. *Pew Internet and American Life Project*, 2009.

[79] Ian Li, Anind Dey, and Jodi Forlizzi. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 557–566. ACM, 2010.

[80] Vivien Lim. The it way of loafing on the job: Cyberloafing, neutralizing and organizational justice. 23:675 – 694, 08 2002.

[81] Ruoyun Lin and Sonja Utz. The emotional responses of browsing facebook: Happiness, envy, and the role of tie strength. *Computers in human behavior*, 52:29–38, 2015.

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

[83] Elizabeth J Lyons, Zakkoyya H Lewis, Brian G Mayrsohn, and Jennifer L Rowland. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. *Journal of medical Internet research*, 16(8):e192, 2014.

[84] Stephen Marche. Is facebook making us lonely. *The Atlantic*, 309(4):60–69, 2012.

[85] Gloria Mark. *Multitasking in the Digital Age.* Morgan and Claypool Publishers, 2015.

[86] Gloria Mark, Daniela Gudith, and Ulrich Klocke. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 107–110. ACM, 2008.

[87] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, Paul Johns, Akane Sano, and Yuliya Lutchyn. Email duration, batching and self-interruption: Patterns of email use on productivity and stress. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1717–1728. ACM, 2016.

[88] Gloria Mark, Yiran Department of Informatics Wang, Melissa Niiya, and Stephanie Reich. Sleep debt in student life: Online attention focus, facebook, and mood. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5517–5528, New York, NY, USA, 2016. ACM.

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

[90] Susan Michie, Maartje M van Stralen, and Robert West. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Science*, 6(1):1, 2011.

[91] Amy Muise, Emily Christofides, and Serge Desmarais. More information than you ever wanted: Does facebook bring out the green-eyed monster of jealousy? *CyberPsychology & behavior*, 12(4):441–444, 2009.

[92] Ashwini Nadkarni and Stefan G Hofmann. Why do people use facebook? *Personality and individual differences*, 52(3):243–249, 2012.

[93] M. Neve, P. J. Morgan, P. R. Jones, and C. E. Collins. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. *Obesity Reviews*, 11(4):306–321.

[94] Mark W Newman, Debra Lauterbach, Sean A Munson, Paul Resnick, and Margaret E Morris. It's not that i don't have problems, i'm just not putting them on facebook: challenges and opportunities in using online social networks for health. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 341–350. ACM, 2011.

[95] Jakob Nielsen. 10 usability heuristics for user interface design. *Nielsen Norman Group*, 1(1), 1995.

[96] Gregory J. Norman, Marion F. Zabinski, Marc A. Adams, Dori E. Rosenberg, Amy L. Yaroch, and Audie A. Atienza. A review of ehealth interventions for physical activity and dietary behavior change. *American Journal of Preventive Medicine*, 33(4):336 – 345.e16, 2007.

[97] Jon Noronha, Eric Hysen, Haoqi Zhang, and Krzysztof Z Gajos. Platemate: crowdsourcing nutritional analysis from food photographs. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 1–12. ACM, 2011.

[98] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile hci. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 919–928. ACM, 2005.

[99] Jeni Paay, Jesper Kjeldskov, Mikael B Skov, Nirojan Srikandarajah, and Umachanger Brinthaparan. Personal counseling on smart phones for smoking cessation. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1427–1432. ACM, 2015.

[100] Pablo Paredes, Ran Gilad-Bachrach, Mary Czerwinski, Asta Roseway, Kael Rowan, and Javier Hernandez. Poptherapy: Coping with stress through pop-culture. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pages 109–117. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.

[101] Loo Geok Pee, Irene M.Y. Woon, and Atreyi Kankanhalli. Explaining non-work-related computing in the workplace: A comparison of alternative models. *Information and Management*, 45(2):120 – 130, 2008.

[102] Laura R Pina, Sang-Wha Sien, Teresa Ward, Jason C Yip, Sean A Munson, James Fogarty, and Julie A Kientz. From personal informatics to family informatics: Understanding family practices around health monitoring. In *CSCW*, pages 2300–2315, 2017.

[103] Josée Poirier and Nathan K Cobb. Social influence as a driver of engagement in a web-based health intervention. *Journal of medical Internet research*, 14(1), 2012.

[104] J O Prochaska and W F Velicer. The transtheoretical model of health behavior change. *American journal of health promotion : AJHP*, 12:38–48, 1997.

[105] John Raacke and Jennifer Bonds-Raacke. Myspace and facebook: Applying the uses and gratifications theory to exploring friend-networking sites. *Cyberpsychology & behavior*, 11(2):169–174, 2008.

[106] Mashfiqui Rabbi, Angela Pfammatter, Mi Zhang, Bonnie Spring, and Tanzeem Choudhury. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. *JMIR mHealth and uHealth*, 3(2):e42–e42, 2014.

[107] Mashfiqui Rabbi, Angela Pfammatter, Mi Zhang, Bonnie Spring, and Tanzeem Choudhury. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: A randomized controlled trial on adults. *JMIR mHealth uHealth*, 3(2):e42, May 2015.

[108] Mashfiqui Rabbi, Angela Pfammatter, Mi Zhang, Bonnie Spring, and Tanzeem Choudhury. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. *JMIR mHealth and uHealth*, 3(2), 2015.

[109] Katharina Reinecke and Krzysztof Z Gajos. Labinthewild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1364–1378. ACM, 2015.

[110] RescueTime. Rescuetime.

[111] Caroline R Richardson, Lorraine R Buis, Adrienne W Janney, David E Goodrich, Ananda Sen, Michael L Hess, Kathleen S Mehari, Laurie A Fortlage, Paul J Resnick, Brian J Zikmund-Fisher, et al. An online community improves adherence in an internet-mediated walking program. part 1: results of a randomized controlled trial. *Journal of medical Internet research*, 12(4), 2010.

[112] K.A. Riekert, J.K. Ockene, and L. Pbert. *The Handbook of Health Behavior Change, 4th Edition*. Springer Publishing Company, 2013.

[113] William T Riley, Katrina J Serrano, Wendy Nilsen, and Audie A Atienza. Mobile and wireless technologies in health behavior and the potential for intensively adaptive interventions. *Current Opinion in Psychology*, 5:67 – 71, 2015. Health behavior.

[114] Tracii Ryan, Andrea Chester, John Reece, and Sophia Xenos. The uses and abuses of facebook: A review of facebook addiction, 2014.

[115] Christina Sagioglou and Tobias Greitemeyer. Facebook's emotional consequences: Why facebook causes a decrease in mood and why people still use it. *Computers in Human Behavior*, 35:359–363, 2014.

[116] William Samuelson and Richard Zeckhauser. Status quo bias in decision making. *Journal of risk and uncertainty*, 1(1):7–59, 1988.

[117] Pedro Sanches, Kristina Höök, Elsa Vaara, Claus Weymann, Markus Bylund, Pedro Ferreira, Nathalie Peira, and Marie Sjölinder. Mind the body!: designing a mobile stress management application encouraging personal reflection. In *Proceedings of the 8th ACM conference on designing interactive systems*, pages 47–56. ACM, 2010.

[118] Sarita Yardi Schoenebeck. 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*, pages 773–782. ACM, 2014.

[119] Muzafer Sherif. A study of some social factors in perception. *Archives of Psychology (Columbia University)*, 1935.

[120] Steven J Sherman. On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology*, 39(2):211, 1980.

[121] Manya Sleeper, Alessandro Acquisti, Lorrie Faith Cranor, Patrick Gage Kelley, Sean A. Munson, and Norman Sadeh. 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*, CSCW '15, pages 1058–1069, New York, NY, USA, 2015. ACM.

[122] Bryant A Stamford, Sharleen Matter, Ronald D Fell, and Paula Papanek. Effects of smoking cessation on weight gain, metabolic rate, caloric consumption, and blood lipids. *The American journal of clinical nutrition*, 43(4):486–494, 1986.

[123] Agnis Stibe and Brian Cugelman. Persuasive backfiring: When behavior change interventions trigger unintended negative outcomes. In Alexander Meschtscherjakov, Boris De Ruyter, Verena Fuchsberger, Martin Murer, and Manfred Tscheligi, editors, *Persuasive Technology*, pages 65–77, Cham, 2016. Springer International Publishing.

[124] Edson C Tandoc, Patrick Ferrucci, and Margaret Duffy. Facebook use, envy, and depression among college students: Is facebooking depressing? *Computers in Human Behavior*, 43:139–146, 2015.

[125] Jih-Hsin Tang, Ming-Chun Chen, Cheng-Ying Yang, Tsai-Yuan Chung, and Yao-An Lee. Personality traits, interpersonal relationships, online social support, and facebook addiction. *Telematics and Informatics*, 33(1):102–108, 2016.

[126] Catalina L Toma and Jeffrey T Hancock. Self-affirmation underlies facebook use. *Personality and Social Psychology Bulletin*, 39(3):321–331, 2013.

[127] Alok Tongaonkar, Shuaifu Dai, Antonio Nucci, and Dawn Song. Understanding mobile app usage patterns using in-app advertisements. In Matthew Roughan and Rocky Chang, editors, *Passive and Active Measurement*, pages 63–72, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[128] Transfusion Media. Stayfocusd.

[129] Ofir Turel, Qinghua He, Gui Xue, Lin Xiao, and Antoine Bechara. Examination of neural systems sub-serving facebook "addiction". *Psychological Reports*, 115(3):675–695, 2014.

[130] Recep Uysal, Seydi Ahmet Satici, and Ahmet Akin. Mediating effect of facebook addiction on the relationship between subjective vitality and subjective happiness. *Psychological reports*, 113(3):948–953, 2013.

[131] Jessica Vitak, Julia Crouse, and Robert LaRose. Personal internet use at work: Understanding cyberslacking. *Computers in Human Behavior*, 27(5):1751 – 1759, 2011. 2009 Fifth International Conference on Intelligent Computing.

[132] L. Thomas Webb, Judith Joseph, Lucy Yardley, and Susan Michie. 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. *J Med Internet Res*, 12(1):e4, Feb 2010.

[133] Robert West, Asha Walia, Natasha Hyder, Lion Shahab, and Susan Michie. Behavior change techniques used by the english stop smoking services and their associations with short-term quit outcomes. *Nicotine & Tobacco Research*, page ntq074, 2010.

[134] McRobbie H Bullen C Borland R Rodgers A Whittaker, R and Y Gu. Mobile phone-based interventions for smoking cessation. *Cochrane Database of Systematic Reviews*, (11), 2012.

[135] Steve Whittaker, Vaiva Kalnikaite, Victoria Hollis, and Andrew Guydish. 'don't waste my time': Use of time information improves focus. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1729–1738. ACM, 2016.

[136] Gabe Zichermann and Christopher Cunningham. *Gamification by design: Implementing game mechanics in web and mobile apps.* " O'Reilly Media, Inc.", 2011.