

Not Now, Ask Later: Users Weaken Their Behavior Change Regimen Over Time, But Expect To Re-Strengthen It Imminently

Geza Kovacs
Lilt, Inc
San Francisco, CA, USA

Zhengxuan Wu
Stanford University
Stanford, CA, USA

Michael S. Bernstein
Stanford University
Stanford, CA, USA

ABSTRACT

How effectively do we adhere to nudges and interventions that help us control our online browsing habits? If we have a temporary lapse and disable the behavior change system, do we later resume our adherence, or has the dam broken? In this paper, we investigate these questions through log analyses of 8,000+ users on HabitLab, a behavior change platform that helps users reduce their time online. We find that, while users typically begin with high-challenge interventions, over time they allow themselves to slip into easier and easier interventions. Despite this, many still expect to return to the harder interventions imminently: they repeatedly choose to be asked to change difficulty again on the next visit, declining to have the system save their preference for easy interventions.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

behavior change; distractions and interruptions

ACM Reference Format:

Geza Kovacs, Zhengxuan Wu, and Michael S. Bernstein. 2021. Not Now, Ask Later: Users Weaken Their Behavior Change Regimen Over Time, But Expect To Re-Strengthen It Imminently. In *CHI '21: ACM CHI Conference on Human Factors in Computing Systems, May 08–13, 2021, Yokohama, Japan*. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3411764.3445695>

1 INTRODUCTION

More people are working with computers [102] and online [32, 71, 98] than ever before, making distractions an ever-present problem [33, 80, 129]. While tied to well-being outcomes [22], social media and other platforms also often lead to self-interruptions that people wish they are better able to control [63, 79]. Many productivity tools have emerged to combat online distractions [62, 77], yet keeping users adhering to interventions remains a challenge [2, 38]. Attrition, where people weaken or give up on their behavior change regimen, can be caused by a number of factors including low perceived intervention effectiveness [63], high perceived intervention

difficulty [45], lack of motivation [6], or a mismatch between the system's interventions and user preferences [63].

In this paper, we seek to understand how people maintain or weaken their behavior change regimens over long periods of time. Are we able to maintain the interventions that we set in place? If we lose the battle, does it happen slowly or suddenly? And once we lose, do we resume our attempt or give up permanently? These questions are critical to the design of behavior change systems, as users may succumb to present-biased choices that are not in line with their long-term goals [73, 120]. In addition to the opportunity to inductively build theory around these questions, there is also a set of practical questions that this research answers: behavior change systems must decide on an appropriate difficulty level [7, 26]: too light a touch, and users might not change their behavior [106], while excessively aggressive interventions may backfire [8, 45, 90]. Knowledge of how user preferences vary over time can help a system identify an appropriate difficulty level for the user in the present moment [1, 11].

In this paper, we study how productivity intervention difficulty preferences change over time, and explore the tradeoffs in terms of time, attrition, and accuracy of asking users about difficulty preferences at various frequencies. We do so by running three studies on the HabitLab platform, an in-the-wild behavior change platform for helping users reduce their time spent online.

An important first question is how users' intervention difficulty preferences evolve over time: what happens to the difficulty levels that users choose over time? How effectively do they stick to their original intended regimen? So, our first study observationally tracks changes in users' choices of intervention difficulty over time. We observe users initially choosing more difficult interventions, and later choosing easier ones, with over half of users eventually keeping the system installed but choosing to have no interventions at all. This result makes clear that user preferences are not static, meaning that any system would need to track changes over time, for example through prompts asking users about their preferences.

Of course, prompting users has attentional and time costs leading to attrition if done too frequently [116]. Thus, in our second study, we observe the costs of prompts in terms of time spent and attrition rates, by randomizing the frequency at which we ask users to choose their desired intervention difficulty levels. We find that excessive prompting significantly increases attrition rates, but that occasional prompting is actually beneficial for retention.

In our third study, we investigate users' future intentions. Specifically, we allow users to not only weaken their interventions, but to save that preference for an hour, a day, or a week. While the most popular intervention level continues to be "No Intervention", the most popular request is to ask again immediately on the next visit. This combination recurs repeatedly, with users continually disabling the system for the current visit but requesting that it try again next

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '21, May 08–13, 2021, Yokohama, Japan

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/10.1145/3411764.3445695>

time rather than stop asking. This snooze button behavior suggests that users remain optimistic about their potential for future behavior change, even if in practice it never materializes.

This paper contributes an analysis of changes in user intervention difficulty preferences in the context of online productivity. We find that users' hope springs eternal: while users choose easier intervention difficulty levels over time and short-term choices can be detrimental to their ability to save time, most users choose to have choices, seemingly expecting to return to more difficult interventions in the near future.

2 RELATED WORK

2.1 Self-Interruptions and Productivity Interventions

Self-interruptions [53] are a widespread occurrence in the workplace [69] and among students [82], which are characterized by users interrupting their work with social media [89], email [81], and recreational web browsing [131]. The relationship of self-interruptions and social media use with well-being is complicated – there can be benefits [22, 105, 107, 127], but excessive social media use can also lead to reduced well-being [21, 24, 65, 128].

A number of sociotechnical approaches have emerged to reduce self-interruptions, including deactivating social network accounts [15], internet addiction bootcamps [67], workplace site filters [46], time trackers [58, 59], as well as various productivity interventions delivered via browser extensions [2, 63, 77, 78], phone applications [57, 62], and chatbots [129].

A challenge in the design of these systems is how much control to give users. In the case of workplace site filters, overly restrictive policies can lower productivity and employee satisfaction [35]. However, if productivity interventions are controlled by users, they can easily be uninstalled or bypassed, and rely on the user remaining committed to continue using them [2, 63]. Thus, productivity tools need to adapt to users [78], which they could do by asking users about their intervention preferences and adapting interventions accordingly.

2.2 Why Lapses Occur: Initial Expectations, Present-Biased Choices, and Self-Control

Sometimes, users choose an intervention – such as deactivating their Facebook account and pledging to never use it again – only to give up and reactivate weeks later [15]. User behavior can lapse for numerous reasons, including declining motivation [17, 84, 104]. Relapse management techniques, which aim to combat such lapses [18, 83, 87, 93, 115], are implemented by some behavior change systems. Examples include “cheat points”, which allow temporary deviations from goals [2], or “streak freezes”, which allow users to maintain a streak without performing the target behavior [51]. Some studies allow users to choose their own intervention [100, 110, 114], but this paper is the first system to study changes in user intervention preferences as reflected by repeated intervention choices over time.

Do users have difficulty sticking to their behavior change regimens because they have unrealistic initial expectations? In dieting contexts, users tend to overestimate their self-control abilities and have unrealistic expectations of their ability to lose weight [124], though this varies by individual [34]. Users likewise underestimate

the amount of time they spend on email and instant messaging when using laptops during lectures [64]. However, while users underestimate the number of times they visit Facebook, they overestimate the time they spend on Facebook [37, 55] and online [9, 113].

Another reason why users struggle to achieve their behavior change goals is that users make short-term choices that conflict with their long-term goals [4]. These manifest themselves as inability to delay gratification, lack of self-control, procrastination, and addiction [73, 120]. Present-biased choices can be attributed to a number of factors – firstly, short-term benefits are more immediate and salient than long-term losses, leading us to discount future outcomes [3, 73, 123]. Additionally, we are often certain of short-term benefits, while long-term effects are less certain, so we discount the uncertain, long-term outcomes [101], or end up considering only a desirable subset of possible outcomes [60, 119]. Optimism can also play a role in present-biased choices, as we are often overly optimistic that we will not suffer from possible negative long-term consequences [56, 136]. Self-control – the ability to resist desires when they conflict with goals – is a key predictor of success [36, 36]. While self-control abilities vary between individuals, situational factors can also influence self-control in the moment [50]. Self-control theories have been used for designing better systems to combat distractions [77].

2.3 Attrition in Behavior Change Systems

Attrition is a major problem faced by behavior change systems [38]. Within the HabitLab system, mismatches between users' intervention difficulty preferences and interventions shown by the system are commonly reported as a reason for uninstalling [63], which motivates us to investigate adapting to user preferences as a means of potentially reducing attrition. That said, attrition in behavior change contexts is influenced by many factors, including lacking time [88], motivation [20], enjoyment of interventions [130], the costs of interventions [52], lacking intention to change [16], intervention novelty [63], unintentionally forgetting about interventions [72], or temporary lapses leading to abandonment [2].

A number of technical approaches help address these issues – for example, adaptive phone and email notifications can help remind users about interventions at the right time [66, 70]. Ambient interventions embedded into routinely used apps, smartwatches, home-screens, or lock screens can encourage engagement during downtime [23, 27, 61, 137]. Gamification approaches such as streaks, points, and giving users cheat points can improve enjoyment and reduce abandonment after temporary lapses [2, 5, 25, 31, 77]. Some systems ask users to make social and financial commitments to encourage them to stick to their goals [44, 99]. Many systems for controlling time online or on phones show interventions automatically during usage, thus reducing attrition via defaults – user inaction will not lead to attrition, as the tools need to be explicitly uninstalled [62, 63, 77, 97].

2.4 Promoting Behavior Change

There are several theoretical frameworks of behavior change [4, 12, 41, 91, 104, 108, 109, 111, 125]. Many of these theories put focal emphasis on the user's commitment to the behavior change regimen [41, 91, 108, 125]. Fogg's B=MAT model, for example,

considers behavior change to occur in the presence of motivation, the ability to take action, and a trigger that prompts people to take action [41]. However, these levels of commitment are challenging to measure, as they depend on the behavior change domain and numerous factors [10, 85], so most approaches rely on self-reporting, which may be unreliable [43]. Some behavior change techniques can provide measurements on related proxies instead [43, 96, 135]. Commitment devices are arrangements where people commit to a plan for achieving a certain behavior goal in the future [19]. They encourage people to stick to their goals by making commitments, such as financial [44, 48] or social [47] commitments. In our work, we draw on the theory of self-shaping: installing software to show interventions, and choosing intervention difficulty levels, can be thought of as a self-shaping commitment device [94].

The field of behavioral economics has developed a number of theoretical frameworks for how to present choices to influence people's choices, known as choice architectures [54, 126]. Defaults are a well-known choice architecture which work by exploiting the status-quo bias [112]. Other widely used choice architectures include limiting the number of choices [29], sorting choices [76], grouping choices [42], and simplifying choice attributes to be more easily interpretable [103, 122]. A number of choice architectures have been developed to combat our bias towards present-biased choices, aversion to uncertainty, and lead us to choices that have better long-term outcomes [54, 123, 133].

Existing studies on self-control and choice architectures have studied contexts where choices only need to be made once or infrequently, and feedback and measurements are often delayed [54]. The context of online productivity provides a superb domain for studying changing user preferences and choices, as we can prompt users multiple times per day, we can vary the frequency of prompting, and the system can provide immediate feedback in response to user choices [63]. This provides us with a more fine-grained lens on how users' preferences change over time.

2.5 Changing Preferences Over Time in Behavior Change Systems

Prior work demonstrates that users will struggle to adhere to their behavior change goals, but the temporal dynamics of this process remain unknown. In this paper, we explore, if users are allowed to choose the difficulty of their interventions, how they navigate the tradeoffs inherent to managing ideal difficulty, and the tradeoffs of different strategies that a behavior change system can pursue to adapt to these changing preferences. This leads us to the following research questions:

RQ1: How do users' intervention difficulty choices change over time? If users' intervention difficulty preferences do not change over time, then behavior change systems can just ask users their preferences during onboarding. However, if they change over time, then behavior change systems may need to continually adapt to users' changing intervention difficulty preferences.

RQ2: Should a behavior change system ask users about their difficulty preferences, and if yes, when and how often? If preferences shift, but the system remains with the user's initial difficulty setting, it could lead to friction or discontinued use. Should the system prompt users about their preferences — or will the act of asking

itself cause attrition? How often should systems prompt users — while more frequent prompting may allow us to more accurately model users' preferences, excessive prompting may have time costs and lead to attrition. We will explore the tradeoffs of prompting frequency with regards to time costs, attrition, and prediction accuracy.

RQ3: Do users prefer to be asked about their intervention difficulty preferences, and if yes, how often do they prefer to be asked? If users' difficulty preferences do not frequently change, we would expect that users would choose to be asked about their difficulty preferences infrequently. However, if users choose to be frequently prompted about their difficulty preferences, yet they keep choosing the same difficulty, this might suggest that users are expecting their future choices to differ from their current choice.

3 EXPERIMENTAL PLATFORM: THE HABITLAB BEHAVIOR CHANGE SYSTEM

To answer these research questions, we conducted three studies on HabitLab [62, 63], an in-the-wild behavior change experimentation platform where users participate in behavior change experiments to help them reduce time online. Users install the browser extension, select sites they wish to reduce time on (*goal sites*), and are shown various productivity interventions when they visit those sites, such as those shown in Figure 1.¹

3.1 Participant Demographics

All participants of the studies in this paper were not recruited or compensated, but were rather all organic installs who discovered HabitLab through sources such as the website, the listing on the Chrome extension store, or press coverage in sources such as Wired or the New York Times. All users whose data we analyzed consented to participate in studies and share their data for research purposes upon installation.

As of this analysis, the HabitLab browser extension has over 12,000 daily active users. According to Google Analytics, 81% of HabitLab users are male, and the most represented age group is 25 to 34. Users are from over 150 countries, and the most represented countries are the USA, Spain, Germany, and Russia. The goal sites they most commonly chose to reduce their time on were Facebook, Youtube, Twitter, Reddit, Gmail, Netflix, and VK.

3.2 Interventions and Difficulty Levels

HabitLab includes interventions to help users reduce their time online, some of which are shown in Figure 1. Some interventions are designed for specific websites such as Facebook, while others are generic and can be used on all sites.

We wished to categorize interventions into difficulty levels. We did so by asking three independent raters (HabitLab users who had been using the platform for over a month) to rate the difficulty level of each intervention as easy, medium, or hard. We opted for a 3-level difficulty categorization, as our studies ask users to choose difficulty levels and we did not want to overwhelm them with too many choices. We took the intervention's difficulty to be the median of its ratings. The Intraclass Correlation Coefficient [118], a statistical measure of inter-rater agreement for ordinal data, was 0.53 —

¹Descriptions of additional interventions can be found in Supplement A. More details about the HabitLab system can be found in [62, 63].

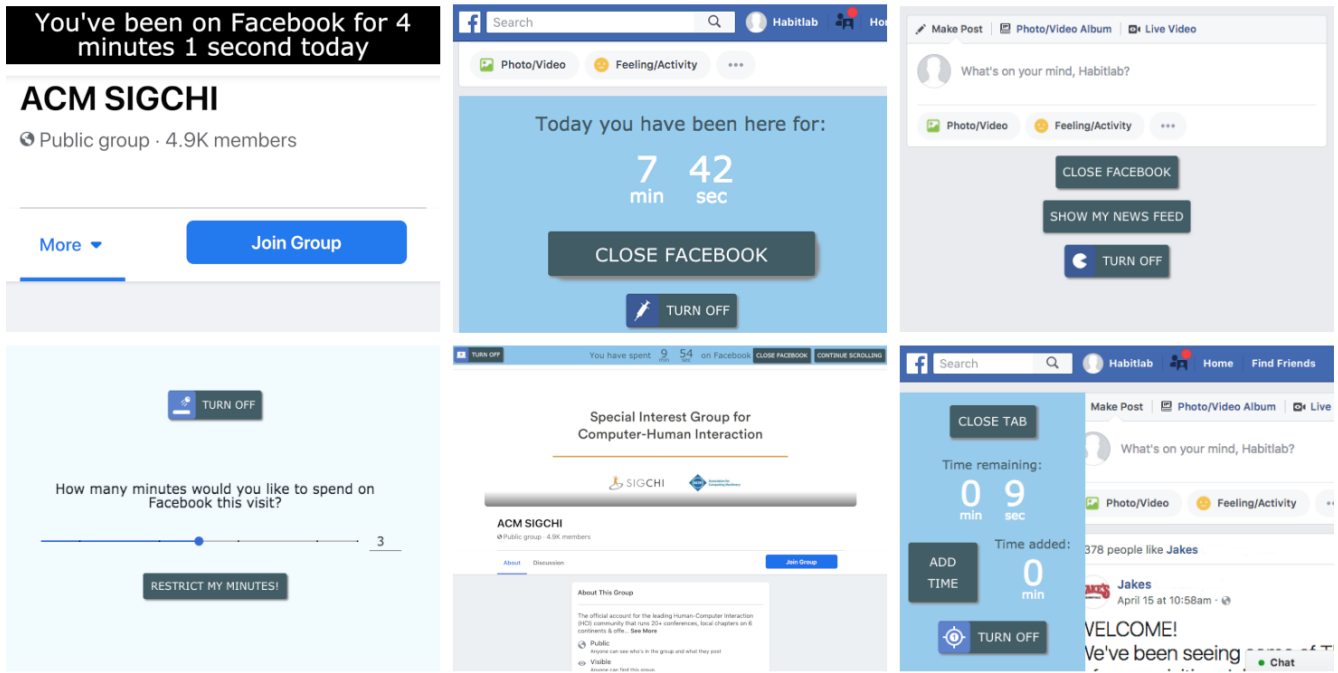


Figure 1: Examples of a few of the many HabitLab interventions available for reducing time on Facebook. From left to right, top to bottom: a timer showing time spent on site at the top of screen (difficulty rated as Easy); a timer injected into the news feed (Easy); requiring the user to opt-in to show the news feed (Medium); requiring the user to set a time limit for how long they will spend this session (Medium); preventing scrolling after a certain number of scrolls until the user clicks a button (Hard); a countdown timer that automatically closes the tab after time elapses (Hard)

indicating that intervention difficulty perceptions may vary between users. Intervention ratings, descriptions, and their effectiveness can be found in Supplement A.

While our definition of intervention difficulty is based on difficulty ratings as opposed to observed effectiveness, interventions rated as more difficult are also more effective. We tested this in a study where on each visit to Facebook, a randomly chosen intervention (or no intervention) is shown. We then measure time spent on Facebook in the presence of that intervention.

A total of 14,139,272 exposure samples were used in this study, from 14,834 users². Our investigation revealed that the most time is spent when there was no intervention (median of 199 seconds per session), followed by easy (185 seconds), medium (161 seconds), and hard (135 seconds) interventions, as shown in Figure 2. There is a significant effect of difficulty on effectiveness according to a Kruskal-Wallis H test ($H=37654$, $p < 0.001$). Differences between pairs of groups are all statistically significant ($p < 0.001$) according to pairwise Mann-Whitney U tests. From this result, we conclude that the difficulty labels capture not only raters' opinions, but also are associated with monotonically increasing time savings when deployed, suggesting that they are in practice more effective.

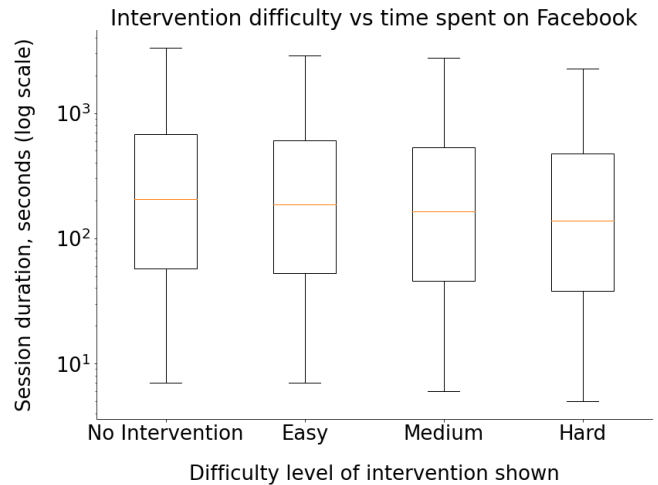


Figure 2: Box plot of Facebook session durations in the presence of interventions. Sessions are significantly shorter in the presence of more difficult interventions. Any intervention difficulty level is more effective than no intervention.

²A more detailed version of this study with Mann-Whitney U-statistic values and per-intervention analyses can be found in Supplement A.

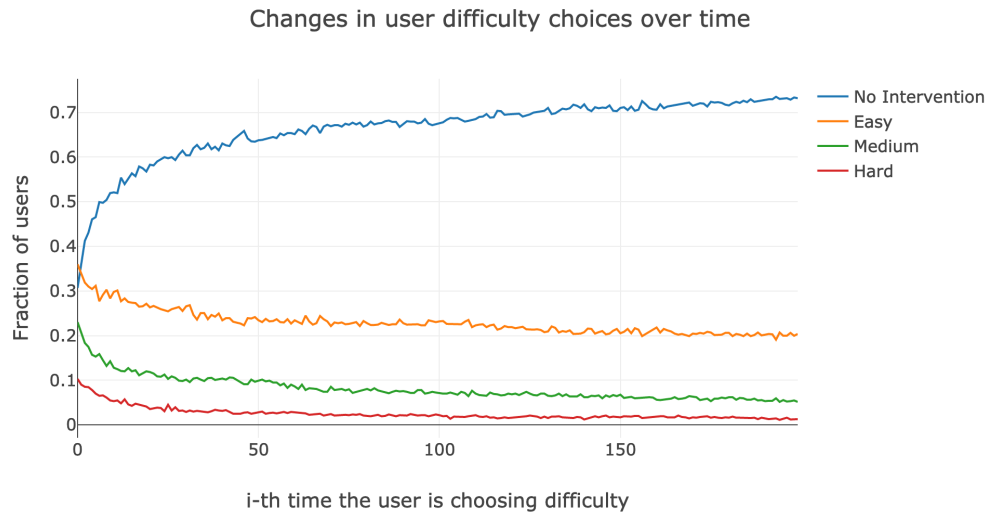


Figure 3: Changes in intervention difficulties chosen by users over time. Users gravitate towards easier interventions over time.

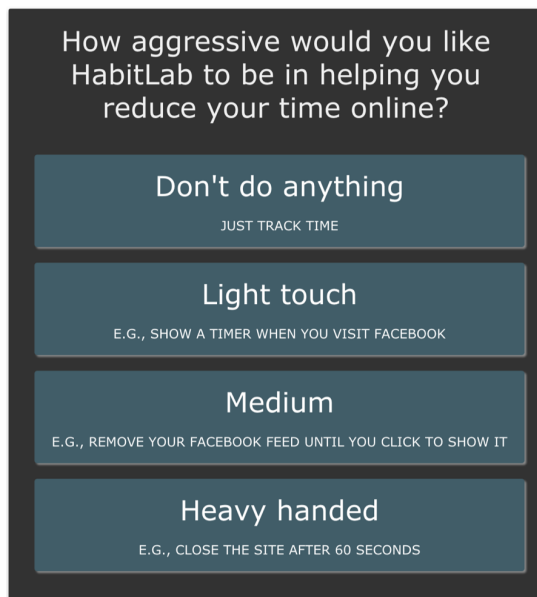


Figure 4: The prompt through which we ask users their preferred intervention difficulty upon visiting a site. A similar prompt is also shown during onboarding.

4 STUDY 1: CHANGES IN INTERVENTION DIFFICULTY CHOICES OVER TIME

In our first study, we seek to understand temporal patterns in how users make choices that balance their commitment to their behavior change regimen against their interest in browsing a goal site. We do so by measuring how users' intervention difficulty preferences change over time, as observed through the intervention difficulty levels they choose on HabitLab. If these preferences are static, then we can just ask about preferences once, and keep them as-is. Many

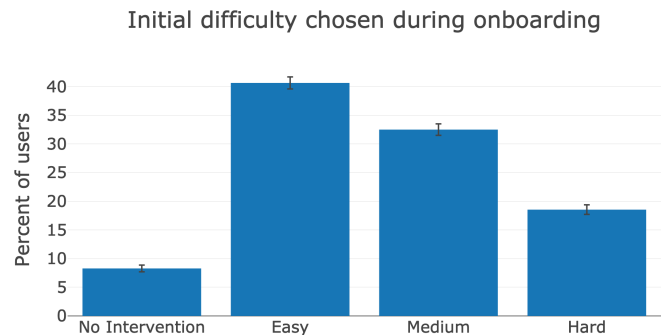


Figure 5: Intervention difficulties chosen by users during onboarding. The most commonly chosen difficulty is easy interventions. Error bars indicate 95% confidence intervals.

behavior change systems implicitly make this assumption, as they only ask the user to state their goals and configure the system during onboarding, and do not later revisit these goals to see whether the user's preferences have changed over time. If intervention difficulty preferences change over time, then understanding the trends will allow our systems to better tailor interventions to users.

4.1 Methodology

When users install HabitLab, we prompt them during onboarding to choose how difficult they would like the interventions to be: No Intervention ("Don't do anything: just track time"), Easy ("Light touch"), Medium ("Medium"), or Hard ("Heavy handed"). Each option is annotated with an example intervention at that difficulty level. Later, as the user continues to use the system, we ask them via a periodic prompt on each visit to a goal site how difficult they would like to have their intervention for that visit, as shown in Figure 4. By tracking changes in the chosen difficulty levels and how they differ from initial preferences indicated during onboarding, we can see how preferences change over time.

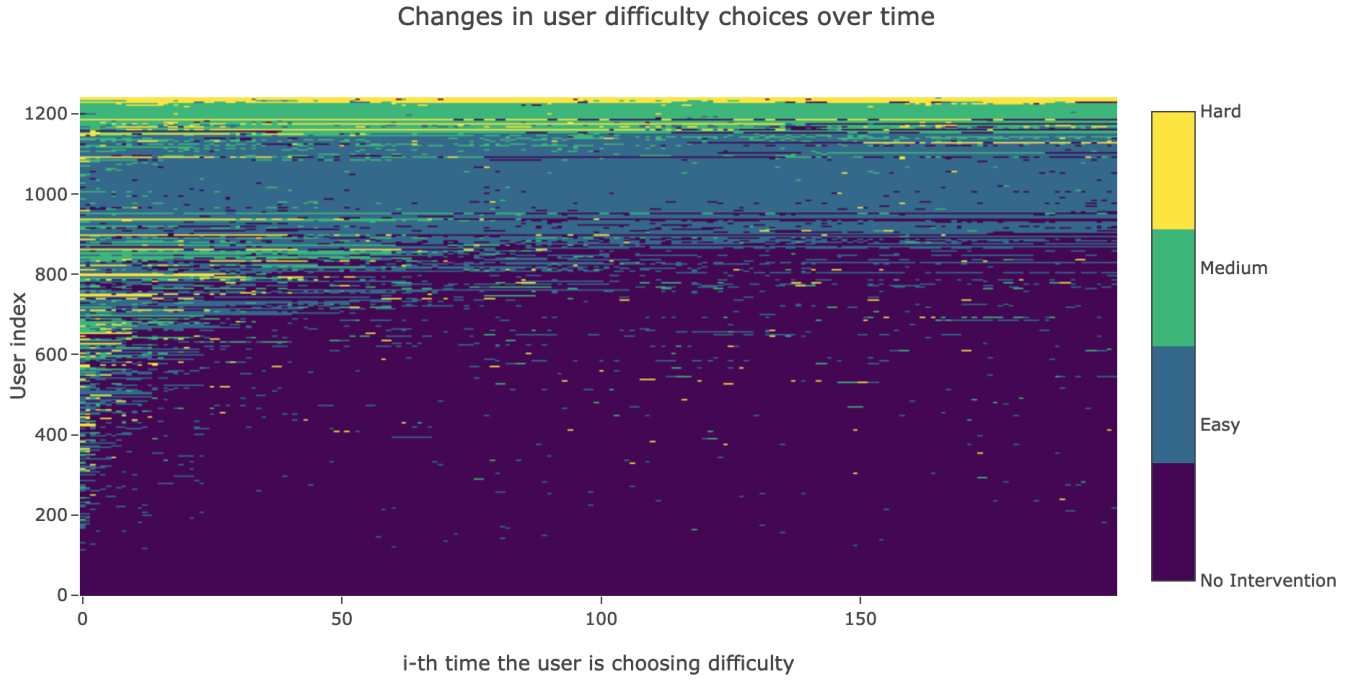


Figure 6: Users begin with more challenging interventions (left) but most slide to easy or no intervention over time (right). The first 200 intervention difficulty choices by each of the 1240 users who made at least 200 intervention difficulty choices. Each user is represented as a row. Time proceeds from left to right. The choice of difficulty is represented by color.

4.2 Results

Responses to the onboarding question of how difficult they would like to have their interventions are shown in Figure 5. The question was answered by 8,372 users. The majority of users desire some form of intervention, with easy interventions being the most frequently chosen option, and no interventions being the least frequently chosen option. A chi-square test indicates there is a significant difference in proportions of responses ($\chi^2 = 2083.4, p < 0.001$). Post-hoc tests on the resulting chi-square contingency table [40, 117] indicate that all pairs of differences are significant ($p < 0.001$).

As we are interested in changes in intervention difficulty preferences over time, we study successive responses to difficulty choice prompts over time. We consider users who have seen and selected an intervention difficulty at least 200 times: a total of 1,240 users during our study period. We visualize this in 2 figures: Figure 3 shows the percent of users who choose each difficulty level at each of the 200 timesteps. Figure 6 visualizes each of the first 200 difficulty choices by each of the user 1,240 users. User preferences initially have a majority of users choosing to have interventions, and many initially go through an exploration phase where they try out different intervention difficulties, which can be seen in Figure 6 as changing colors on the left side. However, over time users choose progressively easier interventions, with 73% of users choosing to be shown no intervention by their 200th visit, as can be seen in Figure 3.

5 STUDY 2: COSTS AND TRADEOFFS OF DIFFICULTY CHOICE PROMPTS

In the first study, we showed that users' intervention difficulty preferences change over time, as indicated by their intervention difficulty choices. Thus, if a behavior change system aims to give users interventions of their desired difficulty, we cannot simply ask about preferences once during onboarding and assume they remains static – the system must continually adapt.

This situation creates challenges for system designers: continually asking users about their preferences may be burdensome and result in attrition. In this section, we measure the costs of asking users to choose a preferred intervention difficulty, and how frequently we need to sample to be able to accurately predict the user's preferred difficulty choices.

5.1 Time costs of difficulty choice prompts

5.1.1 Methodology. We can measure the time costs of difficulty choice prompts by observing the time it takes users to answer the difficulty prompt shown in Figure 4. The time we measure is from when the prompt appears on screen, until the user selects a choice. We consider only sessions where the user actually answers the prompt, as opposed to simply ignoring it.

To determine whether showing the difficulty prompt results in a significant change in duration of visits to goal sites, each time a user visits a goal site the prompt is randomly shown with 50% probability. We measure the overall session lengths – that is, the total time spent from when the user visits a domain until they leave it –

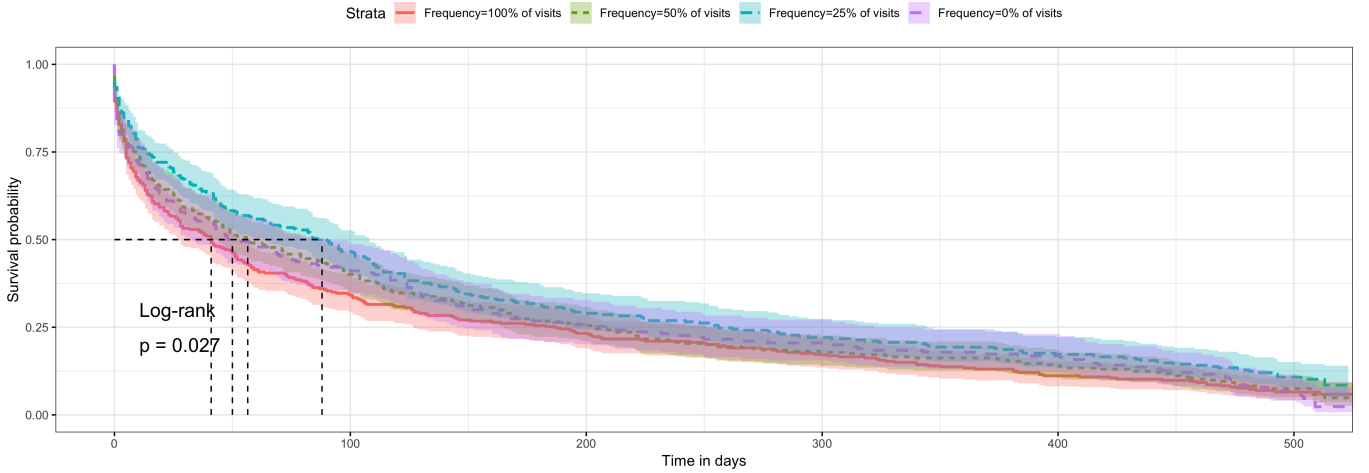


Figure 7: Effects of varying frequencies of difficulty choice prompts on retention rates. Retention is significantly higher when users are asked about their difficulty preferences at a low frequency (25% of visits), compared to being asked on every visit.

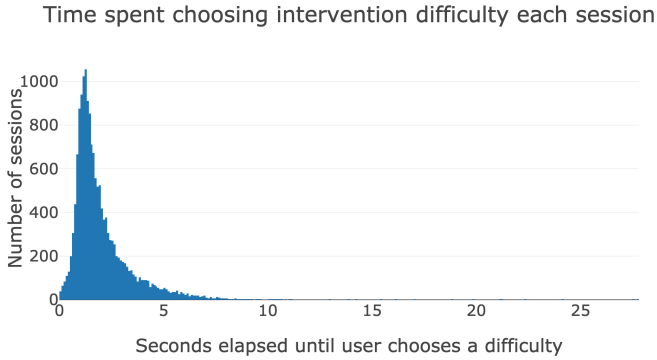


Figure 8: Histogram of time spent answering the difficulty choice prompt. Only sessions where the user made a choice are included. The median is 1.55 seconds.

when the difficulty prompt is shown, vs. when it is not shown. If the difficulty prompt is shown, an intervention of the chosen difficulty is shown; if the difficulty prompt is not shown, an intervention of the most recently chosen difficulty is shown. We then use a linear mixed model [95] to compare log-normalized session lengths in sessions where the prompt was shown to sessions where it was not shown, controlling for the site and user as random effects.

5.1.2 Results. A histogram showing the time spent answering the difficulty prompt is shown in Figure 8. This represents 16,183 responses from 1,831 users. The median of the distribution is 1.55 seconds. We find that there is no significant difference in the duration of sessions when an difficulty prompt is shown, vs not shown. Thus, from the perspective of time spent answering the prompt, our prompts for measuring user difficulty preferences appears to not have major costs. Additionally, the prompt itself does not appear to be influencing time spent on sites.

Frequency	Beta (SE)	Hazard Ratio (95% CI)	p
100% of visits (ref)	-	-	-
50% of visits	-0.09 (0.08)	0.91 (0.78, 1.08)	0.28
25% of visits	-0.25 (0.09)	0.78 (0.66, 0.92)	0.003
0% of visits	-0.07 (0.09)	0.93 (0.78, 1.12)	0.47

Number of events	1,029
Observations	1,108
Concordance	0.533 (SE = 0.01)
Likelihood ratio test	9.43 (df=3, p=0.02)
Wald test	9.24 (df=3, p=0.03)
Log rank test	9.27 (df=3, p=0.03)

Table 1: Effects of varying frequencies of difficulty choice prompts on retention rates. Retention is significantly higher when the prompts are shown 25% of the time, compared to 100% of the time (indicated by the hazard ratio).

5.2 Effects of difficulty choice prompt frequency on retention

The costs of difficulty choice prompts are not restricted to time – they may annoy and distract users, leading to attrition. We received feedback from many users that they had uninstalled HabitLab because they were annoyed by excessive difficulty choice prompts. That said, users often enjoy having their preferences taken into account, and difficulty choice prompts might help users gain a sense of control over the system. Hence, we hypothesized that there may be a tradeoff, with occasional prompting being beneficial, but excessive prompting increasing attrition.

5.2.1 Methodology. We performed an experiment in which we measure the causal effect of changing the interval of difficulty choice prompts on attrition. To do so, we randomly assign users into different conditions according to how frequently we show the difficulty choice prompt. There are 4 conditions: users can be asked 0% of

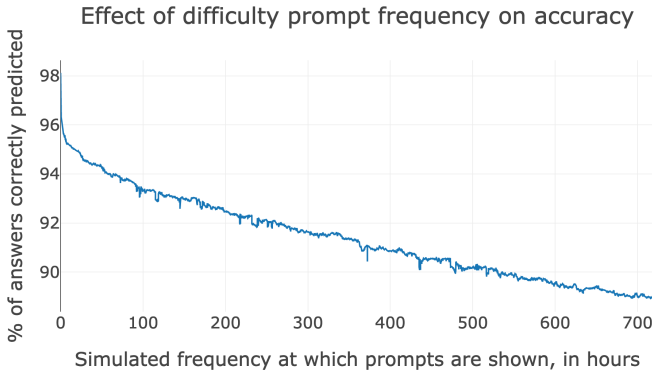


Figure 9: Accuracy of predicting users’ intervention difficulty preferences, given different simulated frequencies of difficulty choice prompts.

visits, 25% of visits, 50% of visits, or 100% of visits. On visits where a prompt was shown, users are shown an intervention of the difficulty they choose in the prompt. If a prompt was not shown, users are shown an intervention of the difficulty they chose the last time they answered the prompt (falling back their difficulty choice during onboarding if they have not yet seen any prompts).³ We ran this experiment with 1108 users over 528 days, and analyzed retention using a Cox hazard regression model [28].

5.2.2 Results. The Cox hazard regression model showing user retention in the different experiment conditions is visualized in Figure 7, while the numeric results are shown in Table 1. We observe via the log-rank test in Table 1 that there is a significant effect ($p < 0.05$) of the prompting frequency on retention. The hazard ratio between the conditions where users are shown the difficulty prompt 25% of the time, and 100% of the time, is below 1 (0.78, see Table 1) and this difference is statistically significant ($p < 0.005$). This means that user retention is significantly higher if users are asked the difficulty question at low frequency (25% of visits), compared to on all visits. We believe it is because showing the prompt occasionally may be beneficial in terms of giving the user a reminder of the system’s presence and granting the user a sense of control. This suggests that asking users about their difficulty preferences at low frequency can strike the right balance of giving users control, while not excessively annoying them.

5.3 Effects of difficulty choice prompt frequency on accuracy

Given that we found that asking users about their difficulty preferences every visit is detrimental to retention, and that asking the same question with lower frequency has higher retention, we would ideally like to use low-frequency difficulty choice prompts to model user preferences. However, there is a tradeoff between frequency and accuracy – asking with lower frequency may lead to lower accuracy. Hence, in this analysis we simulate different frequencies of asking users for their difficulty preference, and observe the accuracy of predicting the actual difficulty chosen by the user. We show results

³We ran a similar experiment where conditions were: each visit / daily / user-chosen intervals, and found similar results; see Supplement C.

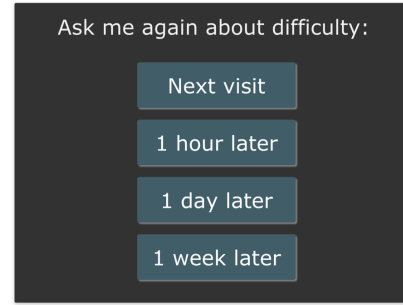


Figure 10: Prompt asking users to choose when to be asked again about difficulty, shown after they choose a difficulty.

for a model that will predict that the user will choose the same difficulty that they chose the last time they saw the difficulty choice prompt. As shown in Figure 9, we can correctly predict the user’s difficulty choice with 96.1% accuracy if we show the difficulty choice prompt at most once per hour, 94.7% accuracy if we ask at most once per day, or 92.8% accuracy if we ask at most once per week. Hence, low frequency prompts are sufficient to accurately predict user intervention difficulty preferences.

6 STUDY 3: USER PREFERENCES FOR DIFFICULTY CHOICE PROMPTS

Our previous study measured behavioral outcomes such as attrition with various frequencies of difficulty choice prompts, but did not take into account user preference at all. We found that retention was improved by asking users about their preferred difficulty at a low frequency, yet there were some users who were sufficiently annoyed by difficulty choice prompts that it led them to uninstall. If given control over prompting frequency, would users want to slide to longer and longer windows of low difficulty or no interventions – as would be consistent with the tendency to regress in difficulty we observed in Study 1? Or would they still choose low-frequency difficulty choice prompts, to maintain their sense of control over intervention difficulty?

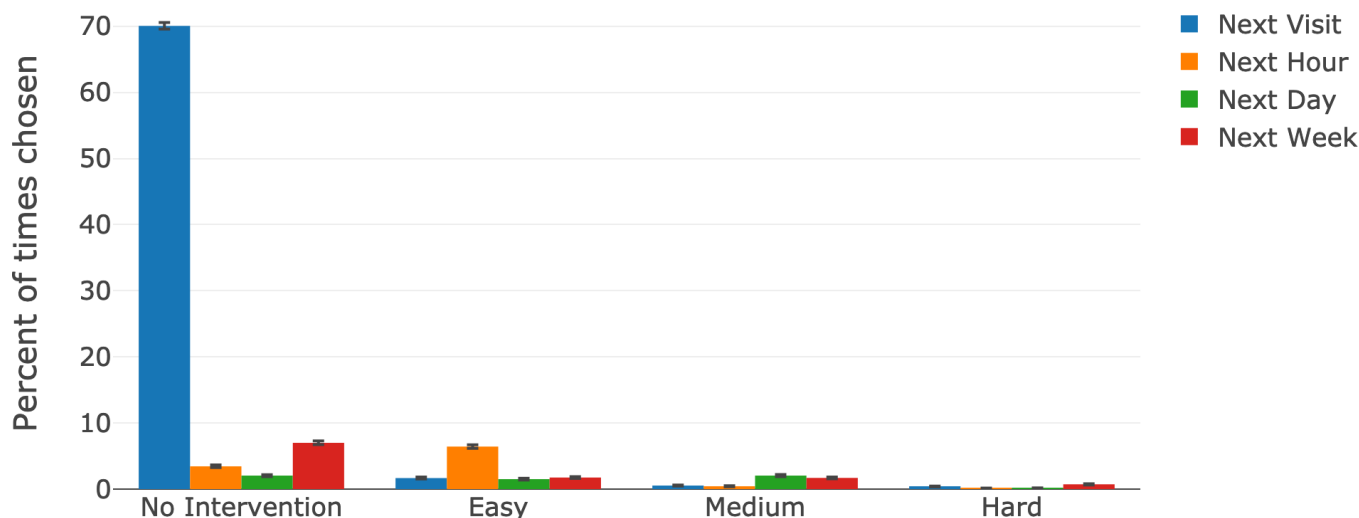
6.1 Methodology

In order to gather this data, we introduced an additional prompt shown immediately after the user selects the intervention difficulty which asks the user when they wish to be prompted again (Figure 10). The interventions that users are shown remain at the difficulty they chose until the next time the prompt is shown. We gathered 31,979 exposure samples from 644 users over the course of 385 days.

6.2 Results

User preferences for when they wish to be asked about intervention difficulty are shown in Figure 12. We observe that users most commonly choose the option to be asked again the next visit. A chi-square test indicates there is a significant difference in proportions of responses ($\chi^2 = 217.79, p < 0.001$). Post-hoc tests on the resulting chi-square contingency table [40, 117] indicate that the difference between Next Hour vs Next Day is not significant ($p > 0.5$), while

Choices for intervention difficulty and when to ask about difficulty again



Intervention difficulty chosen, along with when to ask about difficulty again

Figure 11: Choices for intervention difficulty intersected with when to ask again about difficulty. The most commonly chosen option is to have no intervention this visit, but be asked again on the next visit. Error bars indicate 95% confidence intervals.

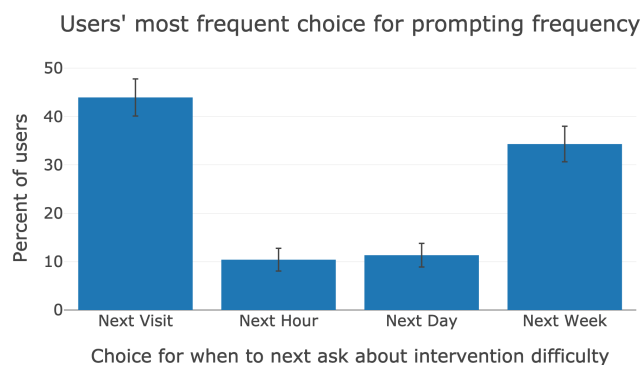
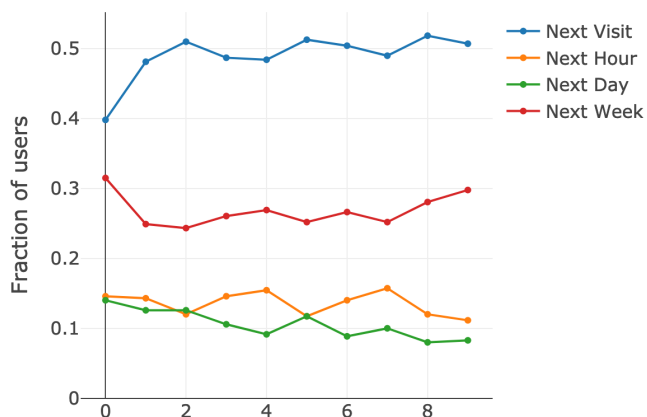


Figure 12: User preferences for frequency of difficulty choice prompts. A plurality (44%) of users most commonly choose to be asked about intervention difficulty on the next visit. Error bars indicate 95% confidence intervals.

all other pairs are significantly different ($p < 0.01$ for Next Visit vs Next Week, and $p < 0.001$ for all others).

To confirm that users' preferences to be asked again about difficulty the next visit is not just a transient phenomenon that goes away over time, we show the change in user choices over time, across the 349 users who made at least 10 choices, in Figure 13. Here, we observe that users' choice of when to be asked again is mostly stable over time, and there is in fact a slight increase in the fraction of users choosing to be asked again next visit over the first 3 visits. This is

Changes in user choices of when to be asked again



i-th time the user is choosing when to be asked next

Figure 13: The first 10 choices of when to be prompted again, among the 349 users who made at least 10 choices.

the opposite trend of we would expect would result from fatigue due to excessive prompting – which would be that users would choose to be shown prompts less frequently over time.⁴

⁴See Supplement D for additional visualizations and analyses over longer periods.

We also visualize the intersection of the intervention difficulty chosen and when the user wishes to be asked again in Figure 11. Here, we find that the most commonly chosen combination is to have no intervention this visit, but to be asked again in the following visit. This result is peculiar, as it seems irrational on the part of users. If users did not want to be bothered by interventions, the logical choice would be to show the intervention difficulty prompt as infrequently as possible – that is, next week. If users wanted to change the intervention difficulty every visit, we would expect to see frequent changes from choosing no intervention to more difficult interventions. However, as we can see in Figure 6 this rarely happens – once a user falls into a pattern of repeatedly choosing to have no intervention, they only occasionally deviate from it.

7 DISCUSSION

This paper is motivated by our observation that users who uninstall HabitLab often cite a mismatch between the difficulty of interventions shown by HabitLab and the difficulty users would prefer as a reason for uninstalling [63]. As a result, we wish to understand changes in users' intervention difficulty preferences over time, and whether it is helpful for behavior change systems to adapt to users' difficulty preferences by prompting them to select intervention difficulty levels. We ran studies on the HabitLab platform to investigate the following research questions:

RQ1: How do users' intervention difficulty choices change over time? We find that user choices of intervention difficulty decline over time. Thus, our behavior change system cannot simply ask users their preferences during onboarding and assume they will remain constant – it needs to continually adapt to changing user preferences.

RQ2: Should a behavior change system ask users about their difficulty preferences, and if yes when and how often? We find that prompting users for their intervention difficulty preference with low frequency has low time costs, and that low-frequency prompting reduces attrition compared to high-frequency prompting. Low-frequency prompting is sufficient to accurately predict user difficulty choices. Thus, a strategy of prompting at low frequency works well for both reducing attrition and adapting to user preferences.

RQ3: Do users prefer to be asked about their intervention difficulty preferences, and if yes, how often do they prefer to be asked? If given a choice of when to be asked again, users will most commonly choose to have no intervention this visit, but to be asked again the next visit. Thus, given users continually ask to be asked again, they appear to expect their future intervention difficulty preferences to change, and do not mind being prompted.

Users are initially optimistic when choosing behavior change interventions – perhaps unrealistically so. We have found that users choose higher difficulty interventions during onboarding than they choose long-term. If we ask users about their desired intervention difficulty later on, it will progressively decline over time.

What are the factors underlying changes in users' intervention difficulty preferences over time? A decline in motivation might occur among HabitLab users over time – in contexts such as volunteering, declines in motivation have been cited as possible reasons for people volunteering less over time [134]. However there are alternative explanations for changing intervention difficulty preferences – it could indicate a shift in priorities, or a decline in commitment to the

goal of reducing time online – perhaps the user installed HabitLab to help them focus during a deadline, and once the deadline has passed they care less about reducing time online. Another alternative is that after repeated exposure, the effectiveness of the intervention declines [63], and users may end up opting for no intervention because they find interventions more distracting than helpful.

Despite users' tendency to choose easier interventions over time, they cling on to hope that they will get back on track. If we ask them when they wish to be asked again about intervention difficulty, by far the most common choice is to have no intervention this visit, but to ask again the next visit.

One explanation for these phenomenon is a combination of users focusing on immediate outcomes when making choices, but attempting to preserve their self-image to be in accordance with their long-term goals when planning about the future. While this behavior seems contradictory, it can be explained from the perspective of reducing cognitive dissonance [39]. Users wish to enjoy the short-term benefits of violating their long-term goals, yet still convince themselves that they will later return to pursuing their long-term goal so they avoid the feeling of having given up. Thus, just like a dieter confronted with temptation may promise themselves that they will only lapse this one time and will stick to their diet in the future – only to repeatedly lapse in the future – HabitLab users may convince themselves that they are only taking a break this one time, and will resume interventions in the future. By retaining the option of resuming in the future, they can continue to reassure themselves that they have not given up on their goal.

An alternative explanation for our result that users consistently choose to have no intervention, but want to be asked again, is that the difficulty choice prompts themselves serve as an intervention that some users feel they need. Perhaps users enjoy the reminder, or they enjoy the sense of choosing, or they think the prompt itself is effective at reducing their time online. As seen by the increase in attrition when users are shown difficulty choice prompts at higher frequencies (Section 5.2), and complaints we received from users about excessive prompting while running this study, this enjoyment of prompting likely does not apply to all users. Additionally, while the presence of prompting can change behavior [14], we did not observe a difference in time spent when the difficulty choice prompt was shown, vs when it was not (Section 5.1), so the prompts themselves do not seem effective as an intervention in this context. However, it is possible that despite annoying some users to the point that they uninstall, other users believe the difficulty choice prompt provides enough value that they constantly ask for it to be shown again.

If we decide we should ask users about their intervention preferences, how often should we do so? In our experiments studying the effects of varying prompting frequency on retention, we found that low prompting frequency results in the highest retention (Section 5.2 and Supplement C). Yet, when choosing when to be asked next about intervention difficulty – which effectively allows users to choose their prompting frequency – they most often choose to be asked again next visit (Section 6), effectively choosing the highest prompting frequency. These results are not necessarily contradictory – people often do not correctly predict their own long-term preferences [74, 121], often prefer retaining the ability to choose [68], the act of choosing itself can influence their future perceptions [75], and there can be individual differences in tolerance for prompting [30].

As for frequency of prompting, if allowing users to choose themselves they split between the two extremes of low and high frequency (Section 6), and preferences do not change over time (Supplement D), but giving users additional control over prompting frequency does not appear to be beneficial for retention or effectiveness (Supplement B and C). This suggests that continually asking users about prompting frequency is not necessary, and keeping prompting at low frequency should be sufficient to avoid attrition.

How might we design behavior change systems for users who wish to retain their self-image as sticking to their long-term goals, but are continually tempted to break them via present-biased choices? It may be the case that the act of choosing itself impairs self-control [132], in which case one option is to remove choices entirely, and not tempt users. For example, just like we could prevent the dieter from encountering desserts and default them to salads, we can simply default users to harder interventions⁵. So long as the interventions are not so far above the user's preferred difficulty level that they drop out, they may continue to stick to it and enjoy the benefits.

Our observation that users weaken their intervention difficulty over time, but appear to hope to re-strengthen it soon, suggests that commitment devices [13, 19, 86, 92] might be worth exploring as potential strategies for keeping users engaged with interventions. Platforms could capitalize on opportunities to try and get users to commit to a slightly more difficult set of interventions, knowing that it might weaken again later. For example, rather than asking the user what difficulty of intervention they would like this visit, we could ask them upfront what difficulty level they will want in the future. By removing the choice, the user's decision is less susceptible to influence from the present. Thus, user choice is not necessarily detrimental — rather, choices should be designed in a way that steers users towards achieving their goals.

7.1 Limitations

Given that perceptions of intervention difficulty may vary depending on the user, one may ask why we chose to have a single categorization of intervention difficulty levels, as opposed to developing a personalized categorization for each user. The latter approach would require users to try out and rate difficulties of all interventions during onboarding before they can start using the system — a lengthy task that would result in our uncompensated, voluntary users not completing onboarding and instead uninstalling. Furthermore, a user's perception of intervention difficulty might change over time, so for truly accurate per-user intervention difficulty ratings we would need additional prompts, which would increase attrition. Per-user intervention difficulties would also complicate statistical analyses, as it would introduce per-user variation in the distributions of randomly chosen easy/medium/hard interventions.

One might ask how the difficulty levels chosen by users may relate to other measures, such as how motivated users are to save time online. Although it is possible that motivation may influence users' choices, it is not necessarily directly observable through users' intervention difficulty choices. We chose to measure users' difficulty

choices rather than asking them about “motivation”, as “motivation” is challenging to define and measure directly — users may not accurately self-report motivation when asked [43].

Our methodology of asking users for their preferred intervention difficulty this visit, and when they would like to be asked again, is somewhat similar to commitment devices as well as experience sampling, but does not match the traditional definition of either. Commitment devices generally ask users to make future commitments [19]. For example, “Would you like to commit to hard interventions for next week?” would be an example of a commitment device in this context. Experience sampling, in turn, differs from our methodology as it generally does not react to the user's choice by immediately presenting an intervention [49]. For example, asking “How difficult did you find that intervention?”, and not acting on the response, would be an example of experience sampling in this context.

8 CONCLUSION

Attrition is a major problem faced by behavior change systems [38, 63], and users commonly report mismatches between the difficulty of interventions shown by the system and users' preferred difficulty as a reason for attrition [63]. In this paper we have explored how users' intervention difficulty preferences change over time, and how behavior change systems can adapt to them. Using prompts on the HabitLab platform, we find that users choose higher intervention difficulty during onboarding, but that their choice of intervention difficulty declines over time. We find that asking users their preferred intervention difficulty at low frequency can both accurately predict the user's preference for intervention difficulty, and can be done at low cost in terms of time and attrition rates.

If we allow users to choose both their desired intervention difficulty as well as when they will be prompted next, they overwhelmingly choose to have no intervention this visit, but to be prompted again next visit. However, users continue to request the system to do nothing, but ask again the next visit, and rarely end up later choosing harder interventions. We believe their choice to have no intervention this visit is driven by present-biased decisions that discount future outcomes, while their choice to be prompted again is driven by a wish to avoid cognitive dissonance and a belief that they will soon get back on track towards achieving their behavior change goals.

Many HCI systems aim to empower users by predicting and understanding users' intentions and preferences, and following them. In the case of behavior change systems, empowering users to achieve their goals requires us to understand users' preferences, while taking into consideration that users may be overly optimistic when initially choosing their behavior change regimen, and may succumb to present-biased choices over time.

REFERENCES

- [1] Marc A Adams, James F Sallis, Gregory J Norman, Melbourne F Hovell, Eric B Hekler, and Elyse Perata. 2013. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. *PLoS one* 8, 12 (2013), e82901.
- [2] Elena Agapie, Daniel Avrahami, and Jennifer Marlow. 2016. Staying the course: System-driven lapse management for supporting behavior change. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1072–1083.
- [3] George Ainslie and Ainslie George. 2001. *Breakdown of will*. Cambridge University Press.
- [4] Icek Ajzen et al. 1991. The theory of planned behavior. *Organizational behavior and human decision processes* 50, 2 (1991), 179–211.

⁵We ran a study testing effects of removing choices and assigning users to default difficulties; see Supplement B.

- [5] Tuomas Alahäivälä and Harri Oinas-Kukkonen. 2016. Understanding persuasion contexts in health gamification: A systematic analysis of gamified health behavior change support systems literature. *International journal of medical informatics* 96 (2016), 62–70.
- [6] Sven Alfonsen, Erik Olsson, and Timo Hursti. 2016. Motivation and treatment credibility predicts dropout, treatment adherence, and clinical outcomes in an internet-based cognitive behavioral relaxation program: a randomized controlled trial. *Journal of medical Internet research* 18, 3 (2016), e52.
- [7] Daniel Almirall, Connie Kasari, Daniel F McCaffrey, and Inbal Nahum-Shani. 2018. Developing optimized adaptive interventions in education. *Journal of research on educational effectiveness* 11, 1 (2018), 27–34.
- [8] James W Anderson, Jingyu Luan, and Lars H Høie. 2004. Structured weight-loss programs: meta-analysis of weight loss at 24 weeks and assessment of effects of intervention intensity. *Advances in Therapy* 21, 2 (2004), 61–75.
- [9] Theo Araujo, Anke Wonneberger, Peter Neijens, and Claes de Vreese. 2017. How much time do you spend online? Understanding and improving the accuracy of self-reported measures of Internet use. *Communication Methods and Measures* 11, 3 (2017), 173–190.
- [10] Cordula Artelt. 2005. Cross-cultural approaches to measuring motivation. *Educational Assessment* 10, 3 (2005), 231–255.
- [11] Elise Baker. 2012. Optimal intervention intensity. *International Journal of Speech-Language Pathology* 14, 5 (2012), 401–409.
- [12] Albert Bandura. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review* 84, 2 (1977), 191.
- [13] Raquel Barata, Paula Castro, and Maria Amélia Martins-Loução. 2017. How to promote conservation behaviours: the combined role of environmental education and commitment. *Environmental Education Research* 23, 9 (2017), 1322–1334.
- [14] Lisa Feldman Barrett and Daniel J Barrett. 2001. An introduction to computerized experience sampling in psychology. *Social Science Computer Review* 19, 2 (2001), 175–185.
- [15] Eric PS Baumer, Phil Adams, Vera D Khovanskaya, Tony C Liao, Madeline E Smith, Victoria Schwanda Sosik, and Kaiton Williams. 2013. Limiting, leaving, and (re) lapsing: an exploration of facebook non-use practices and experiences. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3257–3266.
- [16] Emily Belita and Souraya Sidani. 2015. Attrition in smoking cessation intervention studies: a systematic review. *Canadian Journal of Nursing Research Archive* 47, 4 (2015).
- [17] Mark E Bouton. 2000. A learning theory perspective on lapse, relapse, and the maintenance of behavior change. *Health Psychology* 19, 1S (2000), 57.
- [18] Mark E Bouton. 2014. Why behavior change is difficult to sustain. *Preventive medicine* 68 (2014), 29–36.
- [19] Gharad Bryan, Dean Karlan, and Scott Nelson. 2010. Commitment devices. *Annu. Rev. Econ.* 2, 1 (2010), 671–698.
- [20] Emily Burgess, Peter Hassmén, and Kate L Pumpa. 2017. Determinants of adherence to lifestyle intervention in adults with obesity: a systematic review. *Clinical obesity* 7, 3 (2017), 123–135.
- [21] Moira Burke, Justin Cheng, and Bethany de Gant. 2020. Social Comparison and Facebook: Feedback, Positivity, and Opportunities for Comparison. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [22] Moira Burke and Robert E Kraut. 2016. The relationship between Facebook use and well-being depends on communication type and tie strength. *Journal of computer-mediated communication* 21, 4 (2016), 265–281.
- [23] Carrie J Cai, Philip J Guo, James R Glass, and Robert C Miller. 2015. Wait-learning: Leveraging wait time for second language education. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3701–3710.
- [24] Justin Cheng, Moira Burke, and Elena Goetz Davis. 2019. Understanding perceptions of problematic Facebook use: When people experience negative life impact and a lack of control. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [25] Yu-kai Chou. 2019. *Actionable gamification: Beyond points, badges, and leaderboards*. Packt Publishing Ltd.
- [26] Linda M Collins, Susan A Murphy, and Karen L Bierman. 2004. A conceptual framework for adaptive preventive interventions. *Prevention science* 5, 3 (2004), 185–196.
- [27] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. 2008. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1797–1806.
- [28] David R Cox. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34, 2 (1972), 187–202.
- [29] Henrik Cronqvist and Richard H Thaler. 2004. Design choices in privatized social-security systems: Learning from the Swedish experience. *American Economic Review* 94, 2 (2004), 424–428.
- [30] Mihaly Csikszentmihalyi and Reed Larson. 2014. Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology*. Springer, 35–54.
- [31] Brian Cugelman. 2013. Gamification: what it is and why it matters to digital health behavior change developers. *JMIR serious games* 1, 1 (2013), e3.
- [32] Sara J Czaja. 2020. Setting the Stage: Workplace and Demographic Trends. In *Current and Emerging Trends in Aging and Work*. Springer, 3–11.
- [33] Laura Dabbish, Gloria Mark, and Víctor M González. 2011. Why do i keep interrupting myself?: environment, habit and self-interruption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3127–3130.
- [34] Karina Davidson and Kenneth Prkachin. 1997. Optimism and unrealistic optimism have an interacting impact on health-promoting behavior and knowledge changes. *Personality and social psychology bulletin* 23, 6 (1997), 617–625.
- [35] Pablo Zoghbi Manrique De Lara, Domingo Verano Tacoronte, and Jyh-Ming Ting Ding. 2006. Do current anti-cyberloafing disciplinary practices have a replica in research findings? *Internet Research* (2006).
- [36] Angela L Duckworth, Rachel E White, Alyssa J Matteucci, Annie Shearer, and James J Gross. 2016. A stitch in time: Strategic self-control in high school and college students. *Journal of educational psychology* 108, 3 (2016), 329.
- [37] Sindhu Kiranmai Ernala, Moira Burke, Alex Leavitt, and Nicole B Ellison. 2020. How well do people report time spent on Facebook? An evaluation of established survey questions with recommendations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [38] Gunther Eysenbach. 2005. The law of attrition. *Journal of medical Internet research* 7, 1 (2005), e11.
- [39] Leon Festinger. 1962. *A theory of cognitive dissonance*. Vol. 2. Stanford university press.
- [40] Dustin Fife and Maintainer Dustin Fife. 2017. Package ‘fifer’. *A biostatisticians toolbox for various activities, including plotting, data cleanup, and data analysis* (2017).
- [41] Brian J Fogg. 2009. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*. 1–7.
- [42] Craig R Fox, Rebecca K Ratner, and Daniel S Lieb. 2005. How subjective grouping of options influences choice and allocation: diversification bias and the phenomenon of partition dependence. *Journal of Experimental Psychology: General* 134, 4 (2005), 538.
- [43] Sara M Fulmer and Jan C Rijters. 2009. A review of self-report and alternative approaches in the measurement of student motivation. *Educational Psychology Review* 21, 3 (2009), 219–246.
- [44] Xavier Giné, Dean Karlan, and Jonathan Zinman. 2010. Put your money where your butt is: a commitment contract for smoking cessation. *American Economic Journal: Applied Economics* 2, 4 (2010), 213–35.
- [45] Charles W Given, Barbara A Given, and Bryan W Coyle. 1985. Prediction of patient attrition from experimental behavioral interventions. *Nursing research* (1985).
- [46] Jeremy Glassman, Marilyn Prosch, and Benjamin BM Shao. 2015. To monitor or not to monitor: Effectiveness of a cyberloafing countermeasure. *Information & Management* 52, 2 (2015), 170–182.
- [47] Mary Kay Gugerty. 2007. You can’t save alone: Commitment in rotating savings and credit associations in Kenya. *Economic Development and cultural change* 55, 2 (2007), 251–282.
- [48] Scott D Halpern, David A Asch, and Kevin G Volpp. 2012. Commitment contracts as a way to health. *Bmj* 344 (2012), e522.
- [49] Joel M Hektner, Jennifer A Schmidt, and Mihaly Csikszentmihalyi. 2007. *Experience sampling method: Measuring the quality of everyday life*. Sage.
- [50] Wilhelm Hofmann, Roy F Baumeister, Georg Förster, and Kathleen D Vohs. 2012. Everyday temptations: an experience sampling study of desire, conflict, and self-control. *Journal of personality and social psychology* 102, 6 (2012), 1318.
- [51] Duy Huynh and Hiroyuki Iida. 2017. An analysis of winning streak’s effects in language course of “Duolingo”. *Asia-Pacific Journal of Information Technology and Multimedia* 6, 2 (2017).
- [52] Aurel O Iuga and Maura J McGuire. 2014. Adherence and health care costs. *Risk management and healthcare policy* 7 (2014), 35.
- [53] Jing Jin and Laura A Dabbish. 2009. Self-interruption on the computer: a typology of discretionary task interleaving. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1799–1808.
- [54] Eric J Johnson, Suzanne B Shu, Benedict GC Dellaert, Craig Fox, Daniel G Goldstein, Gerald Häubl, Richard P Larrick, John W Payne, Ellen Peters, David Schkade, et al. 2012. Beyond nudges: Tools of a choice architecture. *Marketing Letters* 23, 2 (2012), 487–504.
- [55] Reynold Junco. 2013. Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior* 29, 3 (2013), 626–631.
- [56] Daniel Kahneman and Dan Lovallo. 1993. Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management science* 39, 1 (1993), 17–31.
- [57] Jaejeung Kim, Chiwoo Cho, and Uichin Lee. 2017. Technology supported behavior restriction for mitigating self-interruptions in multi-device environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous*

- Technologies* 1, 3 (2017), 1–21.
- [58] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, KwonHyun Kim, and Jinwook Seo. 2016. TimeAware: Leveraging framing effects to enhance personal productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 272–283.
 - [59] Young-Ho Kim, Jae Ho Jeon, Bongshin Lee, Eun Kyoung Choe, and Jinwook Seo. 2017. OmniTrack: A flexible self-tracking approach leveraging semi-automated tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–28.
 - [60] Derek J Koehler. 1991. Explanation, imagination, and confidence in judgment. *Psychological bulletin* 110, 3 (1991), 499.
 - [61] Geza Kovacs. 2015. FeedLearn: Using facebook feeds for microlearning. In *Proceedings of the 33rd annual ACM Conference extended abstracts on human factors in computing systems*. 1461–1466.
 - [62] Geza Kovacs, Drew Mylander Gregory, Zilin Ma, Zhengxuan Wu, Golrokh Emami, Jacob Ray, and Michael S Bernstein. 2019. Conservation of Procrastination: Do Productivity Interventions Save Time or Just Redistribute It?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 330.
 - [63] Geza Kovacs, Zhengxuan Wu, and Michael S Bernstein. 2018. Rotating online behavior change interventions increases effectiveness but also increases attrition. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–25.
 - [64] James M Kraushaar and David C Novak. 2019. Examining the affects of student multitasking with laptops during the lecture. *Journal of Information Systems Education* 21, 2 (2019), 11.
 - [65] Ethan Kross, Philippe Verduyn, Emre Demiralp, Jiyoung Park, David Seungjae Lee, Natalie Lin, Holly Shablock, John Jonides, and Oscar Ybarra. 2013. Facebook use predicts declines in subjective well-being in young adults. *PLoS one* 8, 8 (2013), e69841.
 - [66] Florian Künzler. 2019. Context-aware notification management systems for just-in-time adaptive interventions. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 435–436.
 - [67] Alex W Kwee, E Komoru-Venovic, and Janelle L Kwee. 2010. Treatment implications from Etiological and diagnostic considerations of internet Addiction: cautions with the Boot camp Approach. In *Proceedings of the International Conference of e-CASE, Distinguished Paper, CD Format, Singapore*.
 - [68] Lauren A Leotti, Sheena S Iyengar, and Kevin N Ochsner. 2010. Born to choose: The origins and value of the need for control. *Trends in cognitive sciences* 14, 10 (2010), 457–463.
 - [69] Vivien KG Lim and Don JQ Chen. 2012. Cyberloafing at the workplace: gain or drain on work? *Behaviour & Information Technology* 31, 4 (2012), 343–353.
 - [70] Jake Linardon and Matthew Fuller-Tyszkiewicz. 2020. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. *Journal of consulting and clinical psychology* 88, 1 (2020), 1.
 - [71] Kate Lister and Tom Harnish. 2019. Telework and its effects in the United States. In *Telework in the 21st Century*. Edward Elgar Publishing.
 - [72] Jianming Liu, Zhiliang Liu, Hu Ding, and Xiaohong Yang. 2013. Adherence to treatment and influencing factors in a sample of Chinese epilepsy patients. *Epileptic disorders* 15, 3 (2013), 289–294.
 - [73] George Loewenstein and Jon Elster. 1992. *Choice over time*. Russell Sage Foundation.
 - [74] George Loewenstein and David Schkade. 1999. Wouldn't it be nice? Predicting future feelings. *Well-being: The foundations of hedonic psychology* (1999), 85–105.
 - [75] Lennart Luettgau, Claus Tempelmann, Luca Franziska Kaiser, and Gerhard Jocham. 2020. Decisions bias future choices by modifying hippocampal associative memories. *Nature communications* 11, 1 (2020), 1–14.
 - [76] John G Lynch Jr and Dan Ariely. 2000. Wine online: Search costs affect competition on price, quality, and distribution. *Marketing science* 19, 1 (2000), 83–103.
 - [77] Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt. 2019. Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–18.
 - [78] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotko, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2020. 'I Just Want to Hack Myself to Not Get Distracted' Evaluating Design Interventions for Self-Control on Facebook. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.
 - [79] Gloria Mark, Mary Czerwinski, and Shamsi T Iqbal. 2018. Effects of individual differences in blocking workplace distractions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.
 - [80] Gloria Mark, Shamsi Iqbal, Mary Czerwinski, and Paul Johns. 2015. Focused, aroused, but so distractible: Temporal perspectives on multitasking and communications. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 903–916.
 - [81] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, Paul Johns, Akane Sano, and Yuliya Lutchyn. 2016. 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*. 1717–1728.
 - [82] Gloria Mark, Yiran Wang, and Melissa Niiya. 2014. Stress and multitasking in everyday college life: an empirical study of online activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 41–50.
 - [83] G Alan Marlatt. 1996. Taxonomy of high-risk situations for alcohol relapse: evolution and development of a. *Addiction* 91, 12s1 (1996), 37–50.
 - [84] G Alan Marlatt, John S Baer, and Lori A Quigley. 1997. Self-efficacy and addictive behavior. *Self-efficacy in changing societies* (1997), 289–315.
 - [85] Andrew J Martin. 2001. The Student Motivation Scale: A tool for measuring and enhancing motivation. *Australian Journal of guidance and Counselling* 11, 1 (2001), 1–20.
 - [86] Steve J Martin, Suraj Bassi, and Rupert Dunbar-Rees. 2012. Commitments, norms and custard creams—a social influence approach to reducing did not attend (DNAs). *Journal of the Royal Society of Medicine* 105, 3 (2012), 101–104.
 - [87] Robert D Marx. 1982. Relapse prevention for managerial training: A model for maintenance of behavior change. *Academy of management review* 7, 3 (1982), 433–441.
 - [88] Edward McAuley, Kim Poag, Anita Gleason, and Susan Wraith. 1990. Attrition from exercise programs: Attributional and affective perspectives. *Journal of Social Behavior and Personality* 5, 6 (1990), 591.
 - [89] Adrian Meier, Leonard Reinecke, and Christine E Meltzer. 2016. “Facebocrastination”? Predictors of using Facebook for procrastination and its effects on students’ well-being. *Computers in Human Behavior* 64 (2016), 65–76.
 - [90] Kathleen Michael, Andrew P Goldberg, Margarita S Treuth, Jeffrey Beans, Peter Normandt, and Richard F Macko. 2009. Progressive adaptive physical activity in stroke improves balance, gait, and fitness: preliminary results. *Topics in stroke rehabilitation* 16, 2 (2009), 133–139.
 - [91] SI Michie, Marie Johnston, Charles Abraham, R Lawton, D Parker, and A Walker. 2005. Making psychological theory useful for implementing evidence based practice: a consensus approach. *BMJ Quality & Safety* 14, 1 (2005), 26–33.
 - [92] Susan Michie, Michelle Richardson, Marie Johnston, Charles Abraham, Jill Francis, Wendy Hardeman, Martin P Eccles, James Cane, and Caroline E Wood. 2013. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Annals of behavioral medicine* 46, 1 (2013), 81–95.
 - [93] Kathryn R Middleton, Stephen D Anton, and Michal G Perri. 2013. Long-term adherence to health behavior change. *American journal of lifestyle medicine* 7, 6 (2013), 395–404.
 - [94] Neema Moraveji, Ryo Akasaka, Roy Pea, and BJ Fogg. 2011. The role of commitment devices and self-shaping in persuasive technology. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. 1591–1596.
 - [95] John Ashworth Nelder and Robert WM Wedderburn. 1972. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)* 135, 3 (1972), 370–384.
 - [96] Michael Oduor and Harri Oinas-Kukkonen. 2017. Commitment devices as behavior change support systems: a study of users’ perceived competence and continuance intention. In *International Conference on Persuasive Technology*. Springer, 201–213.
 - [97] Fabian Okeke, Michael Sobolev, Nicola Dell, and Deborah Estrin. 2018. Good vibrations: can a digital nudge reduce digital overload?. In *Proceedings of the 20th international conference on human-computer interaction with mobile devices and services*. 1–12.
 - [98] Sabrina Wulff Pabilonia and Victoria Vernon. 2020. Telework and Time Use in the United States. (2020).
 - [99] Katrina Panovich. 2013. Habitbot: A Twitter-based Tool to Help Form and Maintain Habits and Goals.
 - [100] Pablo Paredes, Ran Gilad-Bachrach, Mary Czerwinski, Asta Roseway, Kael Rowan, and Javier Hernandez. 2014. PopTherapy: Coping with stress through pop-culture. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*. ICST (Institute for Computer Sciences, Social-Informatics and ...), 109–117.
 - [101] John W Payne, Namika Sagara, Suzanne B Shu, Kirstin C Appelt, and Eric J Johnson. 2013. Life expectancy as a constructed belief: Evidence of a live-to or die-by framing effect. *Journal of Risk and Uncertainty* 46, 1 (2013), 27–50.
 - [102] Ismael Peña-López et al. 2002. A nation online: How Americans are expanding their use of the Internet. (2002).
 - [103] Ellen Peters, Nathan F Diekmann, Daniel Västfjäll, CK Mertz, Paul Slovic, and Judith H Hibbard. 2009. Bringing meaning to numbers: The impact of evaluative categories on decisions. *Journal of experimental psychology: applied* 15, 3 (2009), 213.
 - [104] James O Prochaska and Wayne F Velicer. 1997. The transtheoretical model of health behavior change. *American journal of health promotion* 12, 1 (1997), 38–48.

- [105] Farzana Quoquab, Zarina Abdul Salam, and Siti Halimah. 2015. Does cyberloafing boost employee productivity?. In *2015 International Symposium on Technology Management and Emerging Technologies (ISTMET)*. IEEE, 119–122.
- [106] Megan Racey, Charlene O'Brien, Sabrina Douglas, Olivia Marquez, Gilly Hendrie, and Genevieve Newton. 2016. Systematic review of school-based interventions to modify dietary behavior: does intervention intensity impact effectiveness? *Journal of School Health* 86, 6 (2016), 452–463.
- [107] Leonard Reinecke, Peter Vorderer, and Katharina Knop. 2014. Entertainment 2.0? The role of intrinsic and extrinsic need satisfaction for the enjoyment of Facebook use. *Journal of Communication* 64, 3 (2014), 417–438.
- [108] RE Rhodes and KS Courneya. 2004. Differentiating motivation and control in the theory of planned behavior. *Psychology, Health & Medicine* 9, 2 (2004), 205–215.
- [109] Craig S Rosen. 2000. Is the sequencing of change processes by stage consistent across health problems? A meta-analysis. *Health psychology* 19, 6 (2000), 593.
- [110] Emily Q Rosenzweig, Judith M Harackiewicz, Stacy J Priniski, Cameron A Hecht, Elizabeth A Canning, Yoi Tibbetts, and Janet S Hyde. 2019. Choose your own intervention: Using choice to enhance the effectiveness of a utility-value intervention. *Motivation Science* 5, 3 (2019), 269.
- [111] Richard M Ryan, Heather Patrick, Edward L Deci, and Geoffrey C Williams. 2008. Facilitating health behaviour change and its maintenance: Interventions based on self-determination theory. *The European health psychologist* 10, 1 (2008), 2–5.
- [112] William Samuelson and Richard Zeckhauser. 1988. Status quo bias in decision making. *Journal of risk and uncertainty* 1, 1 (1988), 7–59.
- [113] Michael Scharkow. 2016. The accuracy of self-reported internet use—A validation study using client log data. *Communication Methods and Measures* 10, 1 (2016), 13–27.
- [114] Stephen M Schueller. 2010. Preferences for positive psychology exercises. *The Journal of Positive Psychology* 5, 3 (2010), 192–203.
- [115] Ralf Schwarzer. 2008. Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied psychology* 57, 1 (2008), 1–29.
- [116] Christie Napa Scollon, Chu-Kim Prieto, and Ed Diener. 2009. Experience sampling: promises and pitfalls, strength and weaknesses. In *Assessing well-being*. Springer, 157–180.
- [117] Donald Sharpe. 2015. Chi-Square Test is Statistically Significant: Now What? *Practical Assessment, Research, and Evaluation* 20, 1 (2015), 8.
- [118] Patrick E Shrout and Joseph L Fleiss. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 86, 2 (1979), 420.
- [119] Suzanne B Shu. 2008. Future-biased search: the quest for the ideal. *Journal of Behavioral Decision Making* 21, 4 (2008), 352–377.
- [120] Suzanne B Shu and Ayelet Gneezy. 2010. Procrastination of enjoyable experiences. *Journal of Marketing Research* 47, 5 (2010), 933–944.
- [121] Itamar Simonson. 1990. The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research* 27, 2 (1990), 150–162.
- [122] Jack B Soll, Ralph L Keeney, and Richard P Larrick. 2013. Consumer misunderstanding of credit card use, payments, and debt: Causes and solutions. *Journal of Public Policy & Marketing* 32, 1 (2013), 66–81.
- [123] Dilip Soman, George Ainslie, Shane Frederick, Xiuping Li, John Lynch, Page Moreau, Andrew Mitchell, Daniel Read, Alan Sawyer, Yaacov Trope, et al. 2005. The psychology of intertemporal discounting: Why are distant events valued differently from proximal ones? *Marketing Letters* 16, 3–4 (2005), 347–360.
- [124] Paul Sparks, Richard Shepherd, Nicole Wieringa, and Nicole Zimmermanns. 1995. Perceived behavioural control, unrealistic optimism and dietary change: An exploratory study. *Appetite* 24, 3 (1995), 243–255.
- [125] Pedro J Teixeira, Marta M Marques, Marlene N Silva, Jennifer Brunet, Joan L Duda, Leen Haerens, Jennifer La Guardia, Magnus Lindwall, Chris Lonsdale, David Markland, et al. 2020. A classification of motivation and behavior change techniques used in self-determination theory-based interventions in health contexts. *Motivation Science* 6, 4 (2020), 438.
- [126] Richard Thaler. 1980. Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization* 1, 1 (1980), 39–60.
- [127] Catalina L Toma and Jeffrey T Hancock. 2013. Self-affirmation underlies Facebook use. *Personality and Social Psychology Bulletin* 39, 3 (2013), 321–331.
- [128] Morten Tromholt. 2016. The Facebook experiment: Quitting Facebook leads to higher levels of well-being. *Cyberpsychology, behavior, and social networking* 19, 11 (2016), 661–666.
- [129] Vincent W-S Tseng, Matthew L Lee, Laurent Denoue, and Daniel Avrahami. 2019. Overcoming Distractions during Transitions from Break to Work using a Conversational Website-Blocking System. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [130] Patty Van Cappellen, Lahnna I Catalino, and Barbara L Fredrickson. 2019. A new micro-intervention to increase the enjoyment and continued practice of meditation. *Emotion* (2019).
- [131] Jessica Vitak, Julia Crouse, and Robert LaRose. 2011. Personal Internet use at work: Understanding cyberslacking. *Computers in Human Behavior* 27, 5 (2011), 1751–1759.
- [132] Kathleen D Vohs, Roy F Baumeister, Brandon J Schmeichel, Jean M Twenge, Noelle M Nelson, and Dianne M Tice. 2014. Making choices impairs subsequent self-control: a limited-resource account of decision making, self-regulation, and active initiative. (2014).
- [133] Elke U Weber, Eric J Johnson, Kerry F Milch, Hannah Chang, Jeffrey C Brod-scholl, and Daniel G Goldstein. 2007. Asymmetric discounting in intertemporal choice: A query-theory account. *Psychological science* 18, 6 (2007), 516–523.
- [134] Galit Ventura Yanay and Niza Yanay. 2008. The decline of motivation?: From commitment to dropping out of volunteering. *Nonprofit management and Leadership* 19, 1 (2008), 65–78.
- [135] Anne Birgitta Yeung. 2004. The octagon model of volunteer motivation: Results of a phenomenological analysis. *Voluntas: International Journal of Voluntary and Nonprofit Organizations* 15, 1 (2004), 21–46.
- [136] Gal Zauberman and John G Lynch Jr. 2005. Resource slack and propensity to discount delayed investments of time versus money. *Journal of Experimental Psychology: General* 134, 1 (2005), 23.
- [137] Huaiyi Zhao, Jie Liu, Jiahao Wu, Kaiyu Yao, and Jin Huang. 2018. Watch-Learning: Using the Smartwatch for Secondary Language Vocabulary Learning. In *Proceedings of the Sixth International Symposium of Chinese CHI*. 108–111.