Sticky Goals: Understanding Goal Commitments for Behavioral Changes in the Wild

Hyunsoo Lee
KAIST
Daejeon, South Korea
hslee90@kaist.ac.kr

Auk Kim*
Kangwon National University
Chucheon, South Korea
kimauk@kangwon.ac.kr

Hwajung Hong†
Seoul National University
Seoul, South Korea
hwajunghong@snu.ac.kr

Uichin Lee‡
KAIST
Daejeon, South Korea
uclee@kaist.ac.kr

ABSTRACT

A commitment device, an attempt to bind oneself for a successful goal achievement, has been used as an effective strategy to promote behavior change. However, little is known about how commitment devices are used in the wild, and what aspects of commitment devices are related to goal achievements. In this paper, we explore a large-scale dataset from stickK, an online behavior change support system that provides both financial and social commitments. We characterize the patterns of behavior change goals (e.g., topics and commitment setting) and then perform a series of multilevel regression analyses on goal achievements. Our results reveal that successful goal achievements are largely dependent on the configuration of financial and social commitment devices, and a mixed commitment setting is considered beneficial. We discuss how our findings could inform the design of effective commitment devices, and how large-scale data can be leveraged to support data-driven goal elicitation and customization.

CCS CONCEPTS

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

KEYWORDS

Commitment Device, Goal-Setting, Online Behavior Change Support Systems, Usage Data Analysis

*Corresponding author for Research Method
†Corresponding author for Design Space and Implications
‡Corresponding author for System Design and Evaluation

1 INTRODUCTION

Prior studies have suggested goal-setting as an effective feature for successful behavior change [27, 82]. Goal-setting has been widely used in personal informatics systems to change diverse behaviors such as physical activity [19, 73], productivity [37, 55], and stress management [60]. According to goal-setting theory, a goal has a higher motivational effect when it is associated with goal commitment, i.e., personal determination of pursuing the goal [58, 64], which has been a central underpinning in goal setting theory since its inception [58].

One popular strategy of enhancing goal commitment is to provide tangible rewards such as cash, raffle, and prize, which have been widely used for diverse behavior change goals such as smoking cessation [34], gym attendance [14], and physical activity [69]. Alternatively, we can associate "restriction" or "penalty" with goal failure (e.g., no refund of a gym membership), which is known as a commitment device [21, 84]. According to behavioral economics literature [84], a commitment device typically refers to "an attempt to enforce people's voluntarily imposed restrictions until they have achieved their goals (e.g., restrict access to TV unless goal achievement), or their voluntarily imposed penalties for failing to accomplish their goals (e.g., forfeit money in case of failure)." There is growing empirical evidence that commitment devices (especially associated with "monetary penalty") were shown to be effective in diverse realms of behavior changes such as mitigating procrastination [4], saving money [93], and increasing exercise [26, 75].

Recently, online behavior change support systems (e.g., stickK, Beeminder) were introduced to help people to pursue their goals by leveraging commitment devices. However, we have little knowledge on how people set their commitment devices for diverse behavior change goals and whether such use of commitment devices is effective in "the real world." Prior studies only investigated the effectiveness of commitment devices in controlled experimental settings, mostly targeting a single behavior change in a small scale. To the best of our knowledge, none of them examined a large-scale goal-setting dataset. While goal-setting strategies have been widely used in personal informatics systems [55, 60, 72], the use of commitment devices for persuasive technology design is still at an early stage of the research; e.g., physical activity [73], productivity [51, 52], and security-related behaviors [32], and the use of commitment device in the wild is largely unexplored. Thus, our work aims to broaden

our knowledge of commitment device design space by answering the following research questions:

- **(RQ1)** What are the patterns of behavior change goals (e.g., topics of target behaviors) and how are commitment devices associated with the goals (i.e., financial and social commitments) that people use in an online behavior change support system?
- **(RQ2)** Among various features of commitment devices, what are the variables related to the successful goal achievements in an online behavior change support system?

To answer these questions, we investigate stickK, an online behavior change support system (https://stickk.com). In stickK, a user can set up a goal by creating a commitment contract— a common form of a commitment device, and managing its progress. For a given goal, a user can make a financial commitment; i.e, the predefined money will be automatically deducted per week from a user's credit card to a chosen recipient (e.g., a charity, an anti-charity, or a friend) if a user fails to achieve a weekly goal. Here, the "anti-charity" refers to a list of polarized organizations, and supporting these organizations may be against a user's values or beliefs such as abortion and political issues, which can possibly increase a user's motivation for achieving behavioral change goals. A social commitment can be configured, by enlisting a "referee" for accountability (i.e., endorsing the user's weekly progress), and a "supporter" for encouragement (i.e., providing psychological support).

We analyzed 75,828 publicly available goals that were posted from November 2013 to February 2018 in stickK and performed qualitative and quantitative analyses of commitment device usage in the wild. Our qualitative results revealed that health-related goals, such as losing weight, had the highest share in the system (70%). In addition, our multilevel regression analyses indicated that successful goal achievement was largely dependent on the configuration of both financial and social commitment devices; e.g., the amount of stake and the recipient type (i.e., friend, charity, and anti-charity), and the existence of supporter. Short-term goals were better achieved than long-term goals.

Based on these findings, we discuss the effect of commitment device mechanisms in stickK and explore possible design opportunities to increase the influence of social commitment devices in mixed commitment systems where financial and social commitment devices are used together as in stickK. We further discuss the possibilities of interpersonal informatics for data-driven and evidence-based reflective goal setting by leveraging large-scale data and existing theories.

The key contributions of the work are twofold. First, as the first large-scale study, our work provides a systematic exploration of commitment device design and examines the key aspects of mixed-commitment design widely used in commercial behavior change support systems. Second, we provide practical design implications on boosting "goal commitment" for designing behavior change support systems, which is an important addition to the body of HCI scholarship.

2 BACKGROUND AND RELATED WORK

In this section, we review existing theories related to commitment devices, behavior change and human decision, and recent HCI studies on commitment devices.

2.1 Theoretical Backgrounds

"Present bias," which refers to people's tendency to discount the value of the later reward (e.g., enjoying dessert now than going to a gym), has long been considered as one major cause for a failure in behavior goals [31, 79]. To tackle attitude-behavior inconsistency, a term "commitment device" has been introduced as an effective strategy. Commitment devices refer to arrangements that an individual makes in the present to control his/her temptation and restrict future behavior by making their choice more costly. Commitment devices can take the form of contingent contracts, pledges, and public announcements. A commonly used form is a commitment contract (i.e. deposit contract), that involves people's voluntary money deposit into an account for a behavior goal and retrieval of money only when the goal is accomplished.

Generally, commitment devices have two basic features. First, a person has to bind his/herself with specific consequences to goal failures. Second, a person has to voluntarily decide to use the devices [11, 84]. These two features increase goal commitment as individuals strive to behave consistently towards their goals because personal responsibility could amplify the adverse effect of goal failures [23, 24, 87]. In order to maximize the effect of commitment devices, previous research has suggested that inclusion of behavior change techniques such as goal-setting, self-monitoring and social support [21].

Commitment devices can be broadly categorized into two: "financial" and "social." Financial commitment devices impose penalties such as monetary loss (e.g., forfeiture of money). Past studies have shown that financial commitment devices had positive influence on diverse behavior goals such as smoking cessation [22, 43], weight loss [21], fitness [86], productivity [40, 49], finance management [48], and stress management [89]. The science behind the effect of financial commitment device is due to "loss aversion," which means that gains and losses are evaluated relative to a reference point and that losses loom larger than gains (e.g., losing $10 is more disappointing than gaining $10) [46].

Social commitment device is designed to primarily cause psychological consequences (e.g., a sense of shame, loss of reputation, and social disapproval) as a result of one's failure. Social commitment device is known to create social accountability, a concept that has been proposed as a key mechanism for supporting the maintenance of certain behaviors [18, 73]. For example, an individual can publicly announce their goals on social media and engage with others to remain accountable for their goals [13, 73]. An underlying mechanism of social commitment device can be explained through cognitive dissonance theory, which illustrates human desire to maintain a continuous and consistent self-image [16, 28]. Studies on social commitment device have also proven its influence on diverse behavior change goals such as health behaviors [41] and household saving [48].

Social commitment device (i.e., accountability) can be incorporated into financial commitment device: for example, in stickK, a

referee (e.g., a close friend or a family member) can check weekly goal progress to determine whether to impose financial penalties. In this mixed commitment scenario, a user's "social" accountability may boost one's motivation for behavior changes along with loss aversion—setting anti-charity donation would further catalyze social accountability.

Along with accountability generated from social commitment device, social support is also known to facilitate behavior change and ample research has confirmed the effectiveness of social support [39, 71]. Social support refers to an individual's needs for affection, approval, belonging and security met by significant others [47]. Although such positive influence of both financial (i.e. economic penalty) and social commitment devices (i.e., accountability) has been reported, commitment devices have been relatively underused compared to other behavior change mechanisms (e.g., financial incentive, social support) [84].

Aside from commitment devices, offering an incentive has been a commonly used mechanism for behavior change goals—a concept that has been central to economics and used across diverse realms of studies to influence behavior. Incentive schemes usually grant a contingent reward (e.g., financial incentive) when a desired outcome is induced [21, 94]. This strategy has been successfully employed to diverse behavior change in experimental settings (e.g., substance abuse [67], smoking cessation [12], physical activity [69], weight loss [66],and dietary behavior [95]). But concerns do exist and tend to center on the actual impact of incentives. It has been reported that some incentives are unlikely to work at a cost-effective level in changing certain complex behaviors [81]. Another caveat is that the immediacy of financial incentives is not appropriate to tackle present bias, which requires temporal restriction or temperance to achieve desired outcome [6].

While these incentive schemes seek to guide people to make better choices and focus on gaining a reward, commitment devices are distinct from incentives [84]. Commitment devices focus on removing and reducing choices, thus increasing the likelihood of one's commitment to a behavior goal [6].

## 2.2 HCI Studies on Commitment Device and Behavior Change

There has been increasing interest in how to support longer-term behavior change in HCI communities as well, mainly focusing on building and designing behavior intervention systems that leverage pervasive context-aware technology [61]. Such systems include habit formation apps [56, 90], digital wellbeing apps [53, 55], or just-in-time reminders to change an individual's health-compromising behaviors (e.g., sedentary lifestyle) [15]. Another side of the studies have focused on building up real-world informatics systems to help people with behavior change and enable self-reflection [25, 62, 85]. Such prior works showed that gathering and inspecting self-relevant data could lead to effective self-reflection, thus possibly leading to positive behavior changes [50]. Thus far, however, the underlying assumption is related to the reflection on one's actual behavior data (e.g., self-tracking data) [29] instead of goal setting itself. To facilitate behavior change, several HCI studies have also attempted to use financial/social commitment devices and other

existing behavior change mechanisms (e.g., financial incentives, social support).

While there are wide applications of financial commitment devices in diverse behavior change domains, HCI studies have paid little attention to financial commitment devices (e.g., financial penalty) for persuasive system design [59, 80]. Recent HCI studies using financial incentives for promoting desirable behaviors have been largely studied in the context of crowd work and micro-task management [10, 70], and user engagement as well as proximal health outcomes [2, 45, 63, 92]. Unlike studies on incentives, strategies exploiting loss aversion or penalty have rather been criticized for being unethical, and it has been advised that ethical ramifications should be deeply considered [30]. Furthermore, there is a concern that nagging feelings of the penalty may drive the users away from the system, leading to an abandonment [17, 20]. As a relatively novel tool for persuasive system design, our work sought to examine how financial commitment devices are set in the wild and what aspects of these devices are closely related to goal achievement.

Similar to financial commitment devices, discussions on whether social accountability can be harnessed as a commitment device are also relatively unaddressed in HCI research, both in volume and depth. Existing studies on social accountability mostly focused on social media (e.g., sharing goals to Facebook friends) [73], and it is less clear how prior studies extend to other scenarios. In contrast, the role of social support or sharing in behavior change support systems or existing social platforms have been actively discussed in HCI studies [39, 73, 76]. Social support is critical for people struggling with health issues because they can receive not only emotional support but also personalized information for behavior changes or maintenance [73]. However, recent studies have shown the side effects of social support and social commitment devices (i.e., accountability). For example, according to recent studies of HCI and CSCW on public announcement of a goal and social accountability, the prospect of social accountability may suppress one from making commitments due to several *barriers* such as impression management and commitment overload [73, 76]. Current body of HCI literature calls for an exploration of leveraging both social support and social commitment devices, and understanding their roles in addition to financial commitment device for the better design of practical behavior change support systems.

Overall, these studies provided the foundations for understanding behavior change support systems and how the use of commitment devices affects goal achievement in HCI research context. However, to the best of our knowledge, none of the prior studies considered the effect of mixed commitment devices (financial and social) with the objective of comparatively understanding each device's effectiveness and how it ultimately affects human behavior. With stickK's mixed commitment design, we seek to evaluate how social commitment devices are related to goal achievement when used along with financial commitment devices. Call for such exploration is corroborated by Lowenstein et al.'s study [65], which demanded that the role of psychological levers such as social commitment devices require further exploration (e.g., level of accountability). Moreover, preliminary studies are confined to a single behavior in an experimental setting, whereas our study attempts to identify how commitment devices affect diverse behavior goals using in-the-wild data.
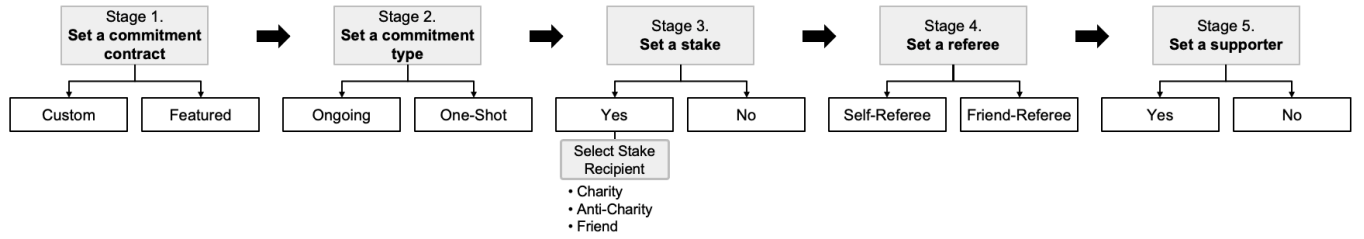
Figure 1: Stages of making a commitment contract in stickK: In stickK, a user can set up his/her goal by creating a commitment contract, and thus each goal can be translated into a commitment contract.
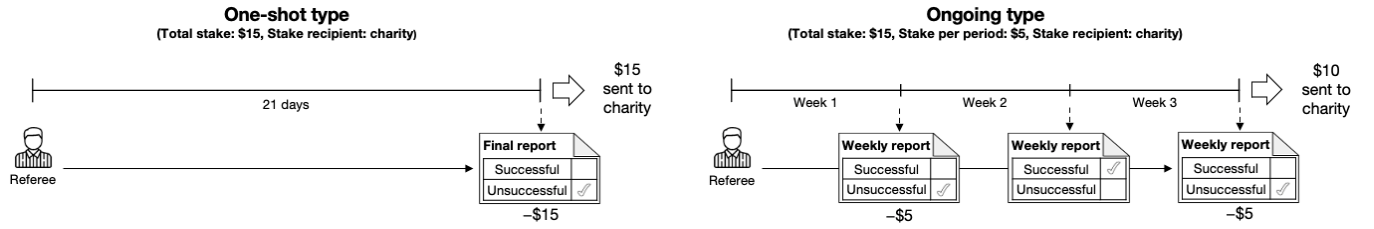
Figure 2: Period Types (Ongoing vs. One-Shot types): Ongoing period types require weekly reports ("Successful" or "Unsuccessful"). In the case of failure, a portion of the total stake amount (i.e., stake per period) will be deducted each week. One-shot period types require a single final report, and the total stake amount will be deducted in the end if a goal fails.

3 MAKING A COMMITMENT CONTRACT IN STICKK

In this work, we consider stickK, a representative online goal-setting platform that provides well-known mechanisms for increasing goal commitment such as goal-setting templates for well-known goals (e.g., weight loss, exercise regularly), weekly self-tracking, and social support via online social goal diaries. Figure 1 illustrates five major steps to create a commitment contract in stickK. In creating a commitment contract, a user can configure financial and social commitment settings for a better goal achievement. In stickK, a user can set up his/her goal by creating a commitment contract, thus each goal can be translated into a commitment contract. For clarity, we call commitment contracts as goals in this paper.

First, a user selects the goal topic between custom and featured goals. stickK provides five featured goals with predefined templates. The five featured goals include "Exercise Regularly," "Lose Weight," "Maintain Weight," "Quit Smoking," and "Race." Such predefined templates provide a structured format which is specific to each goal. In addition to featured goals, stickK allows a user to make a custom goal. A user can describe a goal by entering the title and its content (e.g., sub-goals or activities to perform).

Second, a user is then guided to select a commitment period type with two options, i.e., *ongoing* (recurring weekly reports) or *one-shot* (single report with a deadline). According to stickK's recommendations, an ongoing period type is suitable for the goals that take time for gradual changes and require constant monitoring (e.g., lose weight) whereas a one-shot period type is designed for the tasks that should be done at a particular time (e.g., visiting a doctor, finishing a thesis). If a user selects an ongoing period type, a user has to specify the number of weeks required to achieve a goal. In contrast, when creating a one-shot period type, a user must specify the end date of a commitment contract.

Third, a user can optionally set up a financial commitment by specifying variables of financial commitment device (i.e., an amount of stake, stakes per period and its recipient), which works as follows. In both ongoing and one-shot period types, predefined stakes will be sent to its recipient if a user fails to achieve a goal. In practice, this amount will be automatically charged to the user's registered credit card. Here, note that the deduction of stake occurs differently according to commitment period types. (see Figure 2)

In ongoing period types, a portion of the stake will be deducted from the stake each week and sent to its recipient, whereas one-shot period types require full charge of a stake on the final date. When putting a stake on the line, the minimum and maximum amount of stake per week are $5 and $9,999, respectively. There are three types of recipients: i.e., charity, anti-charity, and friends. The anti-charity refers to a list of organizations that are against one's personal belief or values (e.g., abortion and political issues).

Fourth, a user is required to nominate a referee who will verify the user's weekly progress to remain accountable for making progress on their goals. This accountability-related interaction can be considered as a social commitment device. A user can declare self-checking, or invite a friend to be a referee. A referee's job is to verify a user's weekly report and approve or disapprove of a user's report. If a referee evaluates a user's progress as unsuccessful, or a user self-reports the progress as unsuccessful, or does not submit a progress report within a deadline, a user's weekly progress is marked as a failure.

Lastly, a user is optionally asked to invite friends as supporters who are allowed to make encouraging comments on a user's commitment contract web page (Figure 3). This page also contains
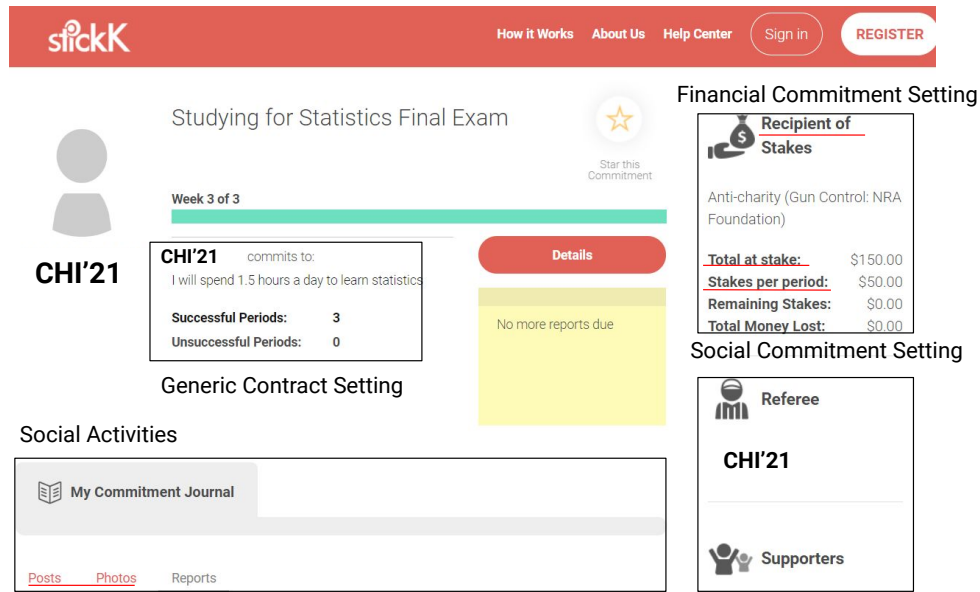
Figure 3: A fictional example of a contract page in stickK. A contract page provides following information: user ID, a commitment title and its description, progress reports (i.e., current week, successful periods, unsuccessful periods), stake information (i.e., recipient, total amount, stake per week, remaining stake, total lost money), referee (for checking), supporters, and commitment journal activities (i.e., posts, photos, reports)

the user's personal commitment journal where a user (or supporters) can post messages or upload photos as well. Through such social activity features given on the system, supporters (as well as a referee) can post supporting messages on a user's commitment journal. As shown in Figure 3, we labeled referee and supporter as "Social Commitment."

4 DATA COLLECTION

We crawled all the publicly available commitment pages in stickK generated from November 2013 to February 2018. The number of collected contracts was 405,648. We removed the 168,351 empty pages that are either deleted pages or private pages, which left us about 237,300 contracts. We removed 97,981 pages with the incomplete contracts that lacked the basic information of a financial commitment device, possibly due to contract cancellation (e.g., a total amount of stake, stake per period). We also removed 63,488 pages with the contracts that were still in progress. As a result, the total number of contracts used for our analysis was 75,828.

We then extracted detailed information from the contract pages using BeautifulSoup, a python library for parsing data out of HTML and XML files. From each page, we extracted the following information: contract number, user ID, title, description of a commitment (word count, character count, average sentence length), stakes per period, total stake, total weeks, successful periods, unsuccessful periods, designated stake recipient (i.e., anti-charity, charity, friend), referee (i.e., self or friend), supporter (i.e., with or without), posts in a commitment journal (i.e., word count, character count, average sentence length), photo upload, and a result of the last four user reports. An example contract page is given in Figure 3.

5 PATTERNS OF TARGET BEHAVIORS AND COMMITMENT DEVICE USAGE

We used a mixed method of qualitative and quantitative analyses to understand the following research question: **"RQ1: What are the patterns of behavior change goals and associated commitment devices that people use in an online behavior change support system?"** In this section, we report on (1) the thematic analysis results of target behaviors' topics from a randomly sampled dataset (Section 5.1); and (2) the descriptive statistics of commitment device usage in the full dataset (Section 5.2).

5.1 Diversity of Target Behaviors

We analyze the topics of target behaviors, which enable us to grasp an understanding of the unique characteristics of behavioral goals generated through online behavior change support systems. We conducted a thematic analysis on 1,000 randomly sampled goals to produce a representative sample of the entire collected dataset.

5.1.1 Methods. Two researchers performed a thematic analysis of the goal description to categorize goals in the sample dataset. We used the approach proposed by Braun and Clarke [9], which consists of 6 phases: familiarization with the data, generating codes, searching for themes, reviewing themes, defining and naming themes, and producing the report. Following the approach, we performed the analysis in a bottom-up, iterative manner.

First, we went through the dataset and generated codes that describe unique themes uncovered from the data. Then, we searched for latent themes in the dataset for further sorting and selected candidates for themes. After several revisions, we defined and named

Table 1: Topics of target behaviors

Category (count)	Sub-Category (count)	Example Goals
Fitness and Wellness (664)	Exercise Regularly & Lose Weight (567)	Lose 10 pounds by next month, To go to the gym 3 times a week
	Food Consumption (61)	No carbs, no sugar!, Drink 8 oz. of water everyday, stay sober
	Smoke Cessation (28)	Not smoke even a single cigarette, Continue to live smoke free, Only smoke if drinking
	Sexual Abstinence (8)	Stop all forms of porn, No Porn. Be a Gentleman. No masturbation
Self-Development (135)	Academic (91)	Get Ph.D. Degree, Studying for Statistics Final Exam, Finish my master thesis
	Hobby Development (25)	Practice Sewing, Post music to Soundcloud
	Career Desire (6)	Get a job at StickK, Launch a successful SEO service
	Business Task (6)	Have all products shipped by Thursday
	Procrastination (4)	Do homework without procrastination
	Language Use (2)	Stop cussing
	Fidgeting Behavior (1)	Stop biting off my nails
Productivity (91)	Time Management (56)	Get up at 6:30 M-F, Early Bedtime, Less time on phone
	Media Consumption (18)	20 min max Facebook per day and Instagram, stop watching tv
	Chores (17)	Weekly household chores, clean underpants 3 times a week!
Journaling and Self-Tracking (47)	Reading & Writing (28)	Write in goal journal for my future, Read for an hour a day
	Self-Tracking (19)	Track calories on my fitness pal, Track Food, Track Sleep
Finance (25)	Budget Control (10)	Control my spending, Minimize petty spending
	Debt (6)	Pay off my student loan, Get out of Debt
	Earn Money (4)	Earn $300k in 2014
	Purchase Behavior (5)	Don't buy what you don't need
Spirituality and Attitude (25)	Meditation (18)	Become more spiritual through meditation
	Religious Activity (6)	Pray for young Christians, Spend time with God
	Negative Mindset (1)	Yell less, be more patient
Social Life (13)	Social Skill (8)	Conquer my habit of avoiding eye contact
	Dating (4)	Dating actively!
	Physical Contact (1)	Not touch XXX (a specific person)

themes. As the last step, we grouped these results into higher-level categories and labeled them to generate representative categories for goals in the dataset. After creating the codebook, two researchers independently coded these 1,000 goals and reached consensus with Cohen's Kappa = 0.97.

5.1.2 Results. We have identified 7 major categories and 26 sub-categories. Here, we report a detailed description of our results and some noteworthy findings from our analysis (see Table 1).

Consistent with prior work [77], we have found that fitness and wellness goals were most common.

Fitness and Wellness describes goals related to promote current health condition, which consists of subcategories related to exercise, weight loss, and food consumption control. Resolutions to deter oneself from unhealthy food consumption or moderation on sensational medium and its following actions were also observed.

Table 2: Summary statistics for the dataset

		Full dataset (SD)	Ongoing (SD)	One-shot (SD)
General information				
Num. of goals		75,828	57,210	18,618
Mean durations in week		15.79 (15.56)	15.79 (15.56)	-
Mean success rate		35.76% (43.73)	29.71% (29.70)	-
Goals with stake				
Num. of goals		25,675	15,600	10,075
Mean stake		$96.04 (483.04)	$100.53 (504.15)	$82.25 (411.18)
Mean stake per period		$27.23(200.71)	$11.02 (62.55)	-
Num. of goals by recipient	Charity	6,749	4,023	2,726
	Anti-charity	14,323	8,413	5,910
	Friend	4,603	3,164	1,439
Mean stake by recipient	Charity	$96.06 (483.08)	$100.55 (504.19)	$82.30 (411.31)
	Anti-charity	$96.04 (483.04)	$100.54 (504.16)	$82.26 (411.18)
	Friend	$96.05 (483.05)	$100.54 (504.15)	$82.32 (411.40)
Num. of goals by referee	Self	17,240	10,032	7,208
	Friend	8,435	5,568	2,867
Num. of goals by supporter	Yes	2,544	1,932	612
	No	23.131	13,668	9,463
Goals without stake				
Num. of goals		50,153	41,610	8,543
Num. of goals by referee	Self	44,104	36,937	7,167
	Friend	6,049	4,673	1,376
Num. of goals by supporter	Yes	2,754	2,229	525
	No	47,399	39,381	8,018

Self-Development refers to goals that reflect aspirations to better oneself. The most frequent sub-category that appeared was related to academic domains which were mainly associated with getting a degree and study a specific subject or a domain. Followed were crafting one's hobby, managing career path, and completing a business task at one's workplace. Despite the relatively small share of this category, it is interesting to note that people use behavior change support systems for goals regarding language use and fidgeting behavior, which are trivial but definitely considered as problematic behaviors generally.

Productivity mostly indicates time management, accounting for almost 62% of this category. Other sub-goals included chores to do and limiting usage of television or social media. *Journaling and Self-Tracking* includes the goals designed to keep track of a certain event or one's actions. Primary goals in this category were writing diaries, tracking exercise, and tracking food intake to stay fit. *Finance* goals describe one's resolution to manage financial status. Such goals included budget control, liquidation of the debt, or earn revenue by a certain period (e.g., paying off loans). Restraining one's impulsive purchase behavior was also found (e.g., not spending any money on coffee).

Spirituality and attitude goals aim to improve one's mental well-being, which includes meditation or religious activity (e.g., reading the bible). Although only one goal was observed, it is worth to highlight that the goal which involves one's attempt to fix a negative

mindset (e.g., practicing patience) was classified as a unique sub-category. *Social Life* goals are related to improving social skills or seeking potential partners.

5.2 Commitment Device Usage in stickK

We overview how users set commitment devices in the online behavior change support system. We summarized the descriptive statistics of the collected data, by quantitatively summarizing the key variables of financial and social commitment devices (i.e., stake amount, stake per week, recipient types, referee types, and supporter selection). We then studied how these variables differ based on the commitment period type: *Ongoing* (recurring) and *One-shot*. Furthermore, we analyzed the featured goals (with preset templates), but overall patterns are very similar, and we did not report the results here. In Table 2, we provide the details about the result from the full dataset with ongoing and one-shot period types. Of the 75,828 goals, 57,210 goals were ongoing period type (75.5%) and 18,618 goals were one-shot type (24.5%).

5.2.1 Descriptive Statistics of Commitment Devices. We present the descriptive statistics related to the goal duration, stake selection, the amount of a stake, stake recipient types, overall success rates, and referee/supporter selection (see Table 2).

Duration: The average duration of goals from the full dataset was 16 weeks in general; to be precise, 16 weeks and 7 weeks from

ongoing and one-shot period types respectively. The distribution of goal duration in ongoing period type ranged from 2 weeks to 176 weeks (M = 16, SD = 16.01), which is slightly over three years and four months. Here, M denotes "mean," and SD indicates "standard deviation." To determine patterns of duration in goals in more detail, we delved into the distribution of goals with ongoing period type within a year. We found that duration of goals is usually set on a monthly basis (i.e., a multiple of four weeks). Relatively short-term goals are prevalent with goals up to 12 weeks (3 months) taking up over 60% of the entire dataset.

Stake Selection: As to optional stake selection, 33.9% (n = 25,675) of the goals had a stake. Considering the commitment period type, 27.3% (n = 15,600) and 54.1% (n = 10,075) of the ongoing (n = 57,210) and one-shot period types (n = 18,618), respectively, had a stake.

Amount of Stake: The average amount of stake was $96, and it ranged from minimum $5 to maximum $9,999.9 (Median = $90). As to the distribution of stake, 86.78% of goals were distributed from $5 to $500, particularly showing high frequencies up to $150. Goals with the ongoing period type were found to put more money than one-shot period type. This is possibly due to the recurring nature of ongoing period type. The average amount of money for ongoing and one-shot period types were $100.53 (SD = 504.15, Median = $120) and$82.25 (SD = 411.18, Median = $50). As to stakes per period, goals generally showed approximately $27.23 (SD = 200.71, Median = $10) per week. A substantially high SD value indicates that the amount of money varies greatly per individual. Note that the range of stake can vary from $5 (min) to $9,999 (max). In the ongoing period type, the average amount of stake per period was $11.02 (SD = 62.55, Median = 10).

Stake Recipient: In stake recipient selection, anti-charity was most popular (n = 14,323), accounting for 55.79% of the goals with a stake. It appears that users prefer anti-charity to boost their goal commitment because they find it difficult to rationalize themselves failing to a goal; if that happens, their money is sent to an organization that offends their personal belief. Followed recipient selections were charity (n = 6,749, 26.29%) and a friend (n = 4,603, 17.93%). The same pattern was also observed in both ongoing and one-shot period types, with anti-charity being the most selected recipient, followed by charity and a friend. Regarding the average amount of stake per recipient, the amount did not differ greatly across commitment period types.

Success Rate: Based on these findings, we also compared the difference in goal achievements between goals with or without a stake. For the comparison, we calculated success rate as follows:

$$Success\ Rate = Successful\ Periods\ /\ Total\ Periods$$

In ongoing period types, a goal is set on a weekly basis. Successful periods denote the number of successful weeks (Figure 3) out of total weeks. Unlike ongoing, one-shot period types have no recurring weekly periods and provide a single period with a deadline. Thereby, the success rate of one-shot period types are either zero ("0" - Not Successful) or one ("1" - Successful).

When comparing the average success rate of goals with stake and those without stake for each user, the success rate of goals with stake was dominantly higher (M = 79.9%, SD = 30.1, Median = 95.8) than goals without stake (M = 9.83%, SD = 2.32 , Median = 0). This contrast indicates that the use of financial commitment device increases the likelihood of successful goal achievements. Given that the number of goals with stake was approximately three times smaller than the number of goals without stake, our finding in the stark contrast between goals with/without stake in terms of performance highlights design challenges in facilitating the use of financial commitment device.

Referee and Supporter: We found that people preferred a self-referee (n = 61,344) to a friend-referee (n = 14,484). Mostly, self-referees accounted for 80.9% in the dataset. When considering the commitment period types, we found that self-referees were chosen for 98.6% and 91.8% of the ongoing and one-shot commitments respectively. As to setting a supporter, a substantial number of goals (93.0%) did not have any supporters.

Social Activities: Along with the aforementioned commitment configuration features, we define photo upload and post upload as social activities. From the data, we found that many of the users didn't take advantage of these variables. Most of the users didn't upload photos (n = 72,860) on their commitment journal to keep track of their goals or certify that they were successful. Although a small proportion of users with photo uploads exists, the users' behavior varied from uploading only 1 to a maximum of 12 photos. Regarding post uploads, approximately 35% of the users posted their progress, with the most active user posting up to 40 posts.

Multiple Goals: We also explored users with multiple goals. Out of the total goals (n = 75,828), 62.4% were distinct users (n = 47,314). 21.7% of the distinct users (n = 10,287) made multiple goals, whereas 78.3% of users (n = 37,027) made a single goal. This result indicates the possibility of stickK's user being only one-time or temporary users to the system. The distribution of multiple goals ranged from minimum 2 to maximum 663 per user and users would generally create 4 goals (SD = 10.03 , Median = 2). An average interval between an initial goal and the next one was approximately 44 days (SD = 118.31, Median = 2).

6 VARIABLES RELATED TO GOAL ACHIEVEMENTS

We explore variables of commitment devices and other software features that affect goal achievements. In this section, we answer RQ2, *"What are the variables related to goal achievements?"*

6.1 Analyses Methods

For the analyses, we conducted a series of mixed model analyses of variance (multilevel models) to examine the relationship between commitment setting variables and successful goal achievement. Since each user may have multiple goals, we include the users as a random effect to control non-independence of the data. In addition, we applied maximum likelihood estimations and grand-mean centering to prevent convergence issues associated with multi-collinearity [54, 78].

For the dependent variable of each analysis, we consistently used the metric called "success rate" to indicate user performance. For a given goal, we calculate the success rate by dividing the number of *Successful Periods* by the number of *Total Periods* (Figure 3). For the independent variables, we considered the following four categories of commitment setting features: generic contract setting

Table 3: Multilevel regression models across commitment period types (Significance: $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$)

Predictor		Ongoing ($R^2 = .782$)			One-shot ($R^2 = .588$)		
		β (SE)	β		β (SE)	β	
<Generic contract setting>							
Length	(week)	-0.002 (0.000)	< .001	***			
	(day)				-0.000 (0.000)	< .001	***
Commitment	upload	-0.013 (0.020)	.51		0.000 (0.000)	.09	
	word count	-0.000 (0.000)	< .001	***	-0.001 (0.001)	.14	
	character count	0.000 (0.000)	.50		-0.001 (0.001)	.25	
	average word length	-0.001 (0.000)	.29		-0.004 (0.016)	.82	
<Social commitment setting>							
Referee		0.038 (0.003)	< .001	***	0.654 (0.008)	< .001	***
Supporter		0.037 (0.004)	< .001	***	-0.046 (0.052)	.37	
<Social activities>							
Photo upload		0.003 (0.005)	.53		0.074 (0.014)	< .001	***
Post	upload	0.096 (0.005)	< .001	***	0.000 (0.000)	.80	
	word count	-0.000 (0.000)	.06		0.000 (0.001)	.86	
	character count	0.000 (0.000)	.10		0.007 (0.003)	< .05	*
	average word length	0.002 (0.001)	.16		0.000 (0.000)	.56	
<Financial commitment setting>							
Total stake		0.000 (0.000)	< .001	***	0.054 (0.008)	< .001	***
Stake per period		0.000 (0.000)	< .001	***	0.000 (0.000)	< .001	***
Recipient	Charity	0.506 (0.004)	< .001	***	0.566 (0.012)	< .001	***
	Anti-charity	0.635 (0.003)	< .001	***	0.532 (0.011)	< .001	***
	Friend	0.530 (0.005)	< .001	***	0.016 (0.012)	.19	

(i.e., goal duration, details of commitment description texts), social commitment setting (i.e., referee, supporter), social activities (i.e., photo, post upload), and financial commitment setting (e.g., stakes per period, total stake, stake recipients). For the measure of goodness-of-fit, we used conditional R^2, which indicates variance explained by both fixed and random effects [74].

6.2 Commitment Period Types

We analyzed and compared significant variables related to goal achievement across commitment period types (i.e., ongoing and one-shot periods).

6.2.1 Ongoing Period Type. We first analyzed which commitment setting factors help users to successfully achieve their goals in ongoing period type.

As shown in Table 3, our results showed that all variables of a financial commitment setting (i.e., stakes per period, total stake, stake recipients), variables of social commitment setting (i.e., referee, supporter), social activities (i.e., post upload) were statistically significant. Goal duration and commitment word counts were also found significant.

The result identified that all the goal-setting variables including financial and social commitment settings were strong predictors of successful commitment. In the financial commitment setting, all the variables showed a strong correlation: i.e., stakes per period ($p < .001$), total stake ($p < .001$), and stake recipient ($p < .001$). Prior HCI studies for behavior change mostly focused on social aspects for behavior change and it is relatively unexplored how one's behavior change is affected when money comes into play.

Our finding highlights the need for further consideration of using money as one of the important design factors. Furthermore, our analysis of featured goals (see Appendix, Table 5) suggests careful consideration of the usage of financial commitment device for a specific behavior (e.g., smoking cessation).

Regarding stake recipients, our analysis shows that selecting anti-charity as the stake recipient is the most effective strategy for success. This is interesting to note, as the finding reflects people's perverse inclination not to have their money sent to an organization that they offend their personal belief or values. Our finding highlights that amplifying a sense of penalty (e.g., anti-charity as stake recipient) to boost commitment is also an important design consideration.

Variables of social commitment setting were also observed to be effective, suggesting that saving one's social face and receiving social support are critical to a successful result: i.e., referee ($p < .001$) and supporter ($p < .001$). Other social activities such as post upload and commitment description also showed positive correlation ($p < .001$). Our finding suggests that enabling social accountability and support along with financial commitment device features can further enhance behavior changes.

Our multilevel regression results also suggest that adopting interaction features such as self-tracking and social-sharing (e.g., photo, post) may boost goal commitment, which is supported by positive correlation between post upload/commitment description and success rate ($p < .001$). As to goal duration, negative correlation was observed. This indicates that the shorter commitment period, the greater likelihood of success. Taken together, these results suggest

Table 4: Multilevel regression model comparing effects of commitment settings between ongoing and one-shot period types (Significance: $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$)

Predictor		Main effects			Interaction effects with Is_One-Shot		
		β (SE)	p		β (SE)	p	
<Generic contract setting>							
Is_One-shot		0.085 (0.050)	.09				
Length (day)		-0.000 (0.000)	< .001	***	-0.000 (0.000)	< .001	***
Commitment upload		-0.004 (0.022)	.85		-0.047 (0.049)	.34	
<Social commitment setting>							
Referee		0.040 (0.003)	< .001	***	0.013 (0.007)	.08	
Supporter		0.039 (0.004)	< .001	***	-0.022 (0.011)	< .05	*
<Social activities>							
Photo upload		-0.002 (0.006)	.77		0.002 (0.014)	.89	
Post upload		0.108 (0.003)	< .001	***	-0.018 (0.006)	< .01	**
<Financial commitment setting>							
Total stake		0.000 (0.000)	< .001	***	0.000 (0.000)	< .05	*
Recipient	Charity	0.527 (0.004)	< .001	***	-0.042 (0.010)	.001	***
	Anti-charity	0.657 (0.004)	< .001	***	-0.050 (0.008)	< .001	***
	Friend	0.548 (0.005)	< .001	***	-0.019 (0.011)	.09	

that it is beneficial to put a stake on the line, set anti-charity as the stake recipient and also invite friends to be a referee.

6.2.2 One-Shot Period Type. Next, we analyzed which commitment setting variables are related to goal achievements in one-shot period type. As shown in Table 3, our analysis result of one-shot period type was almost consistent with that of the ongoing period type. The result showed that most of the variables were statistically significant. All the factors of financial commitment setting (i.e., stakes per period, total stake, stake recipient) showed a strong statistical correlation ($p < .001$). One exceptional finding was found in selecting a friend as the stake recipient, which did not show statistical significance ($p = .19$). Having a friend-referee was also proved to be significant ($p < .001$), whereas supporters did not have a significant impact in one-shot period type. Post upload showed a strong correlation among interaction activities ($p < .001$). Similar to ongoing period type, one-shot period type also showed a negative correlation in goal duration. Such short-term effects of behavior change concurred with past studies [3].

6.2.3 Ongoing Period Type vs. One-Shot Period Type. Finally, we compared how the effects of commitment setting variables on goal achievement varied between ongoing and one-shot period types, using the entire dataset. As shown in Table 4, for the comparison, we additionally included the commitment period type (Is_One-Shot) as a main effect, and its interaction effects with other dependent variables. These additional variables indicate the differences in the effects of commitment setting features between ongoing and one-shot period types. Namely, any statistically significance of these variables indicates that commitment setting features corresponding to the variables are more effective for one-shot period type than ongoing period type. In addition, we removed dependent variables related commitment and post description (i.e., word count, character count, and average word length), which can be considered

as relatively less important variables than other commitment setting variables. We removed these variables to avoid over-fitting (or minimize type I error) since an overly complex model may lead to higher type I error (e.g., singular issues) [7, 68]. In addition, we also removed the stake_per_period feature since the one-shot period type does not have the variable. As shown in Table 4, our result shows that the interaction effects of commitment period type with total stake, stake recipient (charity, anti-charity), goal duration, supporter, and post upload were statistically significant, and indicates these commitment setting variables performed better in one-shot period type. In other words, it is advisable to select one-shot period type when a user goal is suitable for both ongoing and one-shot period types.

6.3 Summary of the Results

We systematically examined the effects of key aspects of mixed-commitment design on goal achievement. Our results can be summarized as follows. First, financial commitment is positively related to goal achievement. Second, amplifying a sense of penalty (e.g., anti-charity as stake recipient) increases goal commitment. Third, social commitment (i.e., accountability) and social support in a mixed commitment setting positively influence goal achievement. Finally, adopting interaction features such as self-tracking and social sharing (e.g., photo, post) may boost goal commitment.

7 DISCUSSION

The purpose of our investigation was to use a data-driven approach to identify aspects of commitment devices that are meaningfully associated with successful goal achievements. By doing so, we found that putting money on the line is more likely to increase the chance of success than proceeding a goal without a stake, highlighting the importance of money as the influential driver in behavior change goals. Positive correlation of social commitment device (i.e.,

friend-referee) and supporters were also observed, indicating that reputational stake and encouragement function as psychological levers in goal achievements. Additionally, social activities such as photo/post upload that may involve other users were also shown to have a positive relationship with goal achievements. In this section, we elucidate our key findings and discuss how these empirical observations can inform future research.

7.1 Financial Commitment Device

7.1.1 Leveraging Loss Aversion. Using a loss aversion strategy as a way to encourage behavioral changes has been used successfully in previous research [44, 46]. In the case of the present study, our multilevel regression analyses showed that the factors related to financial commitment device (i.e., stakes per period, total stake, recipient) dominantly influenced goal achievements in both ongoing and one-shot commitment period types. Our further comparative analysis on three different commitment configurations – *financial commitment* (i.e., stake) only, *social commitments* (i.e., referee and supporter) only, and both *financial and social commitments* also confirmed that money is the potent factor in a goal achievement, with goals configured only with stake showing the highest average success rate (*p* <0.001). Such homogeneity in results indicates that stickK's design choice on loss aversion induces people's resistance against failure, thus increasing commitment to a goal. We also posit that the sunk cost effect may be one of the mechanisms behind the effectiveness of financial commitment device, as people find it hard to leave without achieving their final goals once they have invested funds or effort [5, 42].

7.1.2 Commitment Amplification through Penalty. Furthermore, our analysis identified that the penalty is even more amplified when the recipient was selected to be *anti-charity*, which appears to create an additional level of aversion. Viewing such behavior through Festinger's *Cognitive Dissonance Theory* [28], anti-charity provides extra motivational drives because people find it difficult to rationalize themselves of failing a goal when their money is sent to an organization that offends their personal beliefs. This contrasts with previous studies that altruistic appeal that involves charity reward is more effective in increasing user participation than personal cash incentive [33]. However, these observations also suggest people's strong resistance against loss may have contributed to falsely reported results, which we're unable to find out as users' success relies on self-reports and referee's approval.

Although our results indicate that money is the dominant driver, we do not assert that usage of financial commitment is the direct cause of successful goal achievements. Yardley et al. [96] argues that intervention usage alone cannot be taken as a valid indicator of engagement, but behavioral aspects are also a crucial implication that should be considered when designing for an effective system engagement. Likewise, we posit that latent variables such as one's initial level of motivation or confidence may have affected one's preference in goal-setting (no stake vs. with stake), causing self-selection bias [91]. Further group-wise data analysis to mitigate selection bias is a promising avenue for future work.

Moreover, our results hint that there should be further studies for understanding human behavior and system interaction that

leverage financial commitment device. Future research might envision a design space that explores how people perceive and react to loss aversion under certain goals (e.g., smartphone addiction, work performance) and how such mechanisms are related to people's decision making and behavior change.

7.2 Social Commitment Device and Social Support

Although positive correlation was observed, our regression results indicate that social commitment device and supporters have slightly weaker impact and lower consistency compared to related variables of financial commitment device. As to relative weakness of social commitment device (i.e., friend-referee) and social support, we have several explanations:

7.2.1 Moderation of Financial Commitment. One explanation is that the financial commitment device may have incorporated the "social" accountability, thus increasing the motivation for behavior changes along with loss aversion. As a referee (e.g., a close friend) performs dual duties – 1) checking/encouraging one's weekly goal progress and 2) imposing financial penalties – this may have blurred the role of social commitment. Our comparative analysis on a friend-referee under two different conditions (*with stake* vs. *no stake*) supports such possibility, as it shows a statistical difference in average success rate (82.9% and 25.2% respectively), indicating that the role of a friend-referee may have been affected by the presence of a stake. We also note that differences between the average success rate of self-referee and friend-referee with stake were found statistically insignificant, which suggests that the "social" role of a friend-referee may have been moderated by the financial commitment device. Our postulation can be supported from behavioral economics, where unlike social markets, people's effort in monetary markets is very sensitive to the amount of compensation, and mixed markets more closely resemble monetary than social markets [36].

7.2.2 Psychological Costs of Social Commitment. Another plausible explanation may be due to psychological reasons. While accountability generated from public announcement of a goal is widely recognized as a driving force for social support [1, 72], some people would rather be held accountable by proceeding a goal privately [76]. Existing studies on online social support also argue that people with private personalities may equate sharing their goals and seeking for support with exposing personal vulnerability to others [76], as they fear receiving negative replies and suffer from psychological pressure (e.g., social stigma, impression management, commitment overload) as they share their progress [83, 88]. We posit that such human tendency may have contributed to general preference on self-imposed commitment (i.e., self-referee, no supporter).

Despite aforementioned relative weakness and low selection rates of social commitment device, we were still able to make statistically meaningful observations on the referee's role under 'no stake' condition. From the analysis, we found that friend-referee showed higher average success rate (25.2%) than self-referee (9.46%) even without the involvement of a financial commitment device. Such results and interpretations suggest that there is a need for

increasing the adoption of social commitment device and social support for a mixed commitment setting for commitment boosting.

7.3 Design Implications

Our findings suggest that one's commitment result can differ greatly depending on a commitment contract configuration, highlighting a set of design opportunities in a mixed commitment system. Given such results, it seems reasonable to configure and implement a system that facilitates social support and social commitment device along with financial commitment device. Loewenstein et al. [65] have pointed out that the incentive mechanisms with psychological levers have generally not been compared to the incentives per se, demanding further studies on the effectiveness of mixed commitment mechanisms. Here, we suggest design suggestions to boost one's social engagement in mixed commitment systems.

7.3.1 Reshaping Social Support Networks. As the current state of stickK and other existing platforms take "one-size-fits-all" approaches to supporting social commitment and support, a possible area for design improvement is to allow users to shape and access their own social support networks. Previous discussions from HCI studies have pointed out that different individuals wish to interact with the "appropriate" people to support their goals and select the right channels in terms of goal-setting [73, 76]. Although stickK allows users to designate a referee and invite additional supporters, the efficacy of these functions remain questionable, as they may lack contextual knowledge of a user's personal details and difficulties or complexities involved with a user's goal. One design approach would be a system that automatically suggesting a social cohort-based on personal information of others "like me" [29] with basic information such as similar interests (i.e., target behavior), demographics (e.g., age and gender), performance (i.e., outcome) or sensor data similarities [57].

7.3.2 Increasing Accountability with Nagging Feeling. Harnessing nagging feelings to boost accountability is another design choice worth considering. Our additional analysis on stake recipients shows that financial commitments with stake recipient as anti-charity showed the highest average success rate (87%), followed by friend (77.2%) and charity (75.5%), indicating that leveraging contingent self-extortion as part of an interaction with a system may increase one's resolve. Current landscape of social interaction in stickK provides user-referee/supporter-system triad, but diversifying social actors for interaction may boost a nagging feeling to induce desired behavior change. Ethical concerns do exist, however, because such discomfort may lead to system attrition [17, 20]. Benford et al. [8] argued that deliberately introducing discomfort (emotional or physical discomfort) to interactive experiences can be an important design tool that can realize positive-long term goals. Following this suggestion, leveraging a nagging feeling can be justified if (1) a short-term discomfort leads to longer-term benefits, and (2) an individual's right to choose and withdraw are supported.

7.4 Limitations

Our work is based on a limited dataset due to deleted and private pages. Furthermore, our data cleansing resulted in a substantial loss of contracts because they did not have available information

necessary for our analysis. One caveat of our results is that there may be systematic biases in between those who decided (not) to share their contracts due to privacy reasons. Due to lack of demographic information of users, we could not evaluate the impact of demographic details (e.g., age, gender, socioeconomic status), which may affect the success rates of behavioral change goals [35, 38]. Further studies on similar sites with a comprehensive dataset are required to generalize the findings. Another limitation of this work is that we equally treated all the goals based on their success rate. A user's perception of a success and a goal's difficulty, specificity, and scope may have affected goal achievement. In addition, stickK's focus on a financial commitment device may have led to overrepresentation of the dataset and overestimation of the expected value of financial commitment device usage. However, our descriptive statistics results show that the number of goals without stake (n = 50,153) outweigh the number of goals with a stake (n = 25,675) (see Table 2), indicating that non-financial commitments are also a popular choice by users as much as financial commitments. Lastly, there could be weekly reports falsely marked as a success, which is difficult to control in observational studies.

8 CONCLUSION

Commitment devices offer the potential to help individuals to improve goal commitments by self-restricting their behaviors. Our work builds upon prior experimental studies on commitment device efficacy for behavioral changes by conducting empirical analyses of the commitment contract dataset from stickK, an online behavior change support system. Our thematic analysis and regression analyses revealed the patterns of behavior goals and associated commitment devices and the key predictors of successful goal achievement. We found that successful goal commitments were largely dependent on the configuration of a financial commitment device such as stake amount and recipient type, and there were the positive aspects of nagging feelings with commitment devices for boosting goal commitments. Our results highlighted that in the mixed design of social and financial commitment for behavior change, the role of social support and commitment devices should be carefully considered to improve their influence on goal achievements. As we reflect on implications concerning our findings, we call for further studies on exploring the design space of commitment devices as enabling design tools for behavior changes in diverse domains, ranging from health and wellness to productivity and privacy/security behaviors.

REFERENCES

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

[2] Nabil Alshurafa, Jayalakshmi Jain, Rawan Alharbi, Gleb Iakovlev, Bonnie Spring, and Angela Pfammatter. 2018. Is more always better?: discovering incentivized mHealth intervention engagement related to health behavior trends. *Proceedings*

of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2, 4 (2018), 153.

[3] Jaithri Ananthapavan, A Peterson, and Gary Sacks. 2018. Paying people to lose weight: the effectiveness of financial incentives provided by health insurers for the prevention and management of overweight and obesity–a systematic review. *Obesity reviews* 19, 5 (2018), 605–613.

[4] Dan Ariely and Klaus Wertenbroch. 2002. Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological science* 13, 3 (2002), 219–224.

[5] Hal R Arkes and Catherine Blumer. 1985. The psychology of sunk cost. *Organizational behavior and human decision processes* 35, 1 (1985), 124–140.

[6] Ian Ayres. 2010. *Carrots and sticks: Unlock the power of incentives to get things done.* Bantam.

[7] Douglas Bates, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2015. Parsimonious mixed models. *arXiv preprint arXiv:1506.04967* (2015).

[8] Steve Benford, Chris Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, and Tom Rodden. 2012. Uncomfortable interactions. In *Proceedings of the sigchi conference on human factors in computing systems.* 2005–2014.

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[10] Robin Brewer, Meredith Ringel Morris, and Anne Marie Piper. 2016. "Why would anybody do this?": Older Adults' Understanding of and Experiences with Crowd Work. *Proceedings of CHI 2016* (2016).

[11] Gharad Bryan, Dean Karlan, and Scott Nelson. 2010. Commitment devices. *Annu. Rev. Econ.* 2, 1 (2010), 671–698.

[12] Kate Cahill and Rafael Perera. 2011. Competitions and incentives for smoking cessation. *Cochrane Database of Systematic Reviews* 4 (2011).

[13] Stevie Chancellor, Andrea Hu, and Munmun De Choudhury. 2018. Norms matter: Contrasting social support around behavior change in online weight loss communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–14.

[14] Gary Charness and Uri Gneezy. 2009. Incentives to exercise. *Econometrica* 77, 3 (2009), 909–931.

[15] Woohyeok Choi, Sangkeun Park, Duyeon Kim, Youn-kyung Lim, and Uichin Lee. 2019. Multi-Stage Receptivity Model for Mobile Just-In-Time Health Intervention. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 1–26.

[16] Robert B Cialdini. 2009. *Influence: Science and practice.* Vol. 4. Pearson education Boston.

[17] James Clawson, Jessica A Pater, Andrew D Miller, Elizabeth D Mynatt, and Lena Mamykina. 2015. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM, 647–658.

[18] Sunny Consolvo, Katherine Everitt, Ian Smith, and James A Landay. 2006. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI conference on Human Factors in computing systems.* ACM, 457–466.

[19] Sunny Consolvo, Predrag Klasnja, David W McDonald, and James A Landay. 2009. Goal-setting considerations for persuasive technologies that encourage physical activity. In *Proceedings of the 4th international Conference on Persuasive Technology.* ACM, 8.

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

[21] Nia Coupe, Sarah Peters, Sarah Rhodes, and Sarah Cotterill. 2019. The effect of commitment-making on weight loss and behaviour change in adults with obesity/overweight; a systematic review. *BMC Public Health* 19, 1 (2019), 816.

[22] Jesse Dallery, Steven Meredith, and Irene M Glenn. 2008. A deposit contract method to deliver abstinence reinforcement for cigarette smoking. *Journal of Applied Behavior Analysis* 41, 4 (2008), 609–615.

[23] Gregory C Elliott. 1986. Self-esteem and self-consistency: A theoretical and empirical link between two primary motivations. *Social Psychology Quarterly* (1986), 207–218.

[24] Encyclopedia.com. 2019. Cognitive Consistency Theories. https://www.encyclopedia.com/social-sciences/encyclopedias-almanacs-transcripts-and-maps/cognitive-consistency-theories.

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

[26] Leonard H Epstein, Rena R Wing, J Kevin Thompson, and William Griffin. 1980. Attendance and fitness in aerobics exercise: The effects of contract and lottery procedures. *Behavior Modification* 4, 4 (1980), 465–479.

[27] Paul A Estabrooks, Candace C Nelson, Stanley Xu, Diane King, Elizabeth A Bayliss, Bridget Gaglio, Paul A Nutting, and Russell E Glasgow. 2005. The frequency and behavioral outcomes of goal choices in the self-management of diabetes. *The Diabetes Educator* 31, 3 (2005), 391–400.

[28] Leon Festinger. 1962. *A theory of cognitive dissonance.* Vol. 2. Stanford university press.

[29] Clayton Feustel, Shyamak Aggarwal, Bongshin Lee, and Lauren Wilcox. 2018. People Like Me: Designing for Reflection on Aggregate Cohort Data in Personal Informatics Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 107.

[30] BJ Fogg, Gregory Cueller, and David Danielson. 2007. Motivating, influencing, and persuading users: An introduction to captology. In *The human-computer interaction handbook.* CRC Press, 159–172.

[31] Shane Frederick, George Loewenstein, and Ted O'donoghue. 2002. Time discounting and time preference: A critical review. *Journal of economic literature* 40, 2 (2002), 351–401.

[32] Alisa Frik, Nathan Malkin, Marian Harbach, Eyal Peer, and Serge Egelman. 2019. A Promise Is A Promise: The Effect of Commitment Devices on Computer Security Intentions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* ACM, 604.

[33] David H Furse and David W Stewart. 1982. Monetary incentives versus promised contribution to charity: New evidence on mail survey response. *Journal of Marketing Research* 19, 3 (1982), 375–380.

[34] 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.

[35] Stephanie J Hanrahan and Ester Cerin. 2009. Gender, level of participation, and type of sport: Differences in achievement goal orientation and attributional style. *Journal of science and medicine in sport* 12, 4 (2009), 508–512.

[36] James Heyman and Dan Ariely. 2004. Effort for payment: A tale of two markets. *Psychological science* 15, 11 (2004), 787–793.

[37] Alexis Hiniker, Sungsoo Ray Hong, Tadayoshi Kohno, and Julie A Kientz. 2016. MyTime: designing and evaluating an intervention for smartphone non-use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, 4746–4757.

[38] John R Hollenbeck and Arthur P Brief. 1987. The effects of individual differences and goal origin on goal setting and performance. *Organizational Behavior and Human Decision Processes* 40, 3 (1987), 392–414.

[39] Hwajung Hong, Eric Gilbert, Gregory D Abowd, and Rosa I Arriaga. 2015. In-group questions and out-group answers: crowdsourcing daily living advice for individuals with autism. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, 777–786.

[40] Tanjim Hossain and John A List. 2012. The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science* 58, 12 (2012), 2151–2167.

[41] Kevin O Hwang, Allison J Ottenbacher, Angela P Green, M Roseann Cannon-Diehl, Oneka Richardson, Elmer V Bernstam, and Eric J Thomas. 2010. Social support in an Internet weight loss community. *International journal of medical informatics* 79, 1 (2010), 5–13.

[42] Kazushi Ikeda and Michael S Bernstein. 2016. Pay it backward: Per-task payments on crowdsourcing platforms reduce productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, 4111–4121.

[43] Robert W Jeffery, Jean L Forster, Simone A French, Steven H Kelder, Harry A Lando, Paul G McGovern, David R Jacobs Jr, and Judith E Baxter. 1993. The Healthy Worker Project: a work-site intervention for weight control and smoking cessation. *American journal of public health* 83, 3 (1993), 395–401.

[44] Leslie K John, George Loewenstein, Andrea B Troxel, Laurie Norton, Jennifer E Fassbender, and Kevin G Volpp. 2011. Financial incentives for extended weight loss: a randomized, controlled trial. *Journal of general internal medicine* 26, 6 (2011), 621–626.

[45] Gyuwon Jung, Jio Oh, Youjin Jung, Juho Sun, Ha-Kyung Kong, and Uichin Lee. 2021. "Good Enough!": Flexible Goal Achievement with Margin-based Outcome Evaluation. In *In CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21).* ACM, New York, NY, USA.

[46] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives* 5, 1 (1991), 193–206.

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

[48] Felipe Kast, Stephan Meier, and Dina Pomeranz. 2012. *Under-savers anonymous: Evidence on self-help groups and peer pressure as a savings commitment device.* Technical Report. National Bureau of Economic Research.

[49] Supreet Kaur, Michael Kremer, and Sendhil Mullainathan. 2015. Self-control at work. *Journal of Political Economy* 123, 6 (2015), 1227–1277.

[50] Elisabeth T Kersten-van Dijk, Joyce HDM Westerink, Femke Beute, and Wijnand A IJsselsteijn. 2017. Personal informatics, self-insight, and behavior change: A critical review of current literature. *Human–Computer Interaction* 32, 5-6 (2017), 268–296.

[51] Inyeop Kim, Hwarang Goh, Nematjon Narziev, Youngtae Noh, and Uichin Lee. 2020. Understanding User Contexts and Coping Strategies for Context-Aware Phone Distraction Management System Design. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 134 (Dec. 2020), 33 pages. https:

//doi.org/10.1145/3432213

[52] Inyeop Kim, Uichin Lee, and Narae Cha. 2018. Exploring Context-aware Proactive Blocking for Distraction Management. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. ACM, 1202–1207.

[53] Jaejeung Kim, Hayoung Jung, Minsam Ko, and Uichin Lee. 2019. GoalKeeper: Exploring Interaction Lockout Mechanisms for Regulating Smartphone Use. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–29.

[54] Rex B Kline. 2015. *Principles and practice of structural equation modeling*. Guilford publications.

[55] Minsam Ko, Seungwoo Choi, Subin Yang, Joonwon Lee, and Uichin Lee. 2015. FamiLync: facilitating participatory parental mediation of adolescents' smartphone use. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 867–878.

[56] 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.

[57] Nicholas D Lane, Ye Xu, Hong Lu, Shaohan Hu, Tanzeem Choudhury, Andrew T Campbell, and Feng Zhao. 2011. Enabling large-scale human activity inference on smartphones using community similarity networks (csn). In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 355–364.

[58] Gary P Latham and Edwin A Locke. 1991. Self-regulation through goal setting. *Organizational behavior and human decision processes* 50, 2 (1991), 212–247.

[59] Hyunsoo Lee, Uichin Lee, and Hwajung Hong. 2019. Commitment devices in online behavior change support systems. In *Proceedings of Asian CHI Symposium 2019: Emerging HCI Research Collection*. 105–113.

[60] Kwangyoung Lee and Hwajung Hong. 2018. MindNavigator: Exploring the Stress and Self-Interventions for Mental Wellness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 572.

[61] Uichin Lee, Kyungsik Han, Hyunsung Cho, Kyong-Mee Chung, Hwajung Hong, Sung-Ju Lee, Youngtae Noh, Sooyoung Park, and John M. Carroll. 2019. Intelligent positive computing with mobile, wearable, and IoT devices: Literature review and research directions. *Ad Hoc Networks* 83 (2019), 8–24.

[62] Ian Li, Anind K Dey, and Jodi Forlizzi. 2011. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 405–414.

[63] Brian Y Lim, Judy Kay, and Weilong Liu. 2019. How Does a Nation Walk?: Interpreting Large-Scale Step Count Activity with Weekly Streak Patterns. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 57.

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

[65] George Loewenstein, David A Asch, and Kevin G Volpp. 2013. Behavioral economics holds potential to deliver better results for patients, insurers, and employers. *Health Affairs* 32, 7 (2013), 1244–1250.

[66] Charles Fergus Lowe, Pauline J Horne, Katy Tapper, Michael Bowdery, and Christine Egerton. 2004. Effects of a peer modelling and rewards-based intervention to increase fruit and vegetable consumption in children. *European journal of clinical nutrition* 58, 3 (2004), 510.

[67] Jennifer Plebani Lussier, Sarah H Heil, Joan A Mongeon, Gary J Badger, and Stephen T Higgins. 2006. A meta-analysis of voucher-based reinforcement therapy for substance use disorders. *Addiction* 101, 2 (2006), 192–203.

[68] Hannes Matuschek, Reinhold Kliegl, Shravan Vasishth, Harald Baayen, and Douglas Bates. 2017. Balancing Type I error and power in linear mixed models. *Journal of Memory and Language* 94 (2017), 305–315.

[69] Marc S Mitchell and Paul I Oh. 2016. Framing financial incentives to increase physical activity among overweight and obese adults. *Annals of internal medicine* 165, 8 (2016), 599–600.

[70] Tanushree Mitra, Clayton J Hutto, and Eric Gilbert. 2015. Comparing person-and process-centric strategies for obtaining quality data on amazon mechanical turk. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1345–1354.

[71] David Mohr, Pim Cuijpers, and Kenneth Lehman. 2011. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. *Journal of medical Internet research* 13, 1 (2011), e30.

[72] Sean A Munson and Sunny Consolvo. 2012. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *2012 6th international conference on pervasive computing technologies for healthcare (PervasiveHealth) and workshops*. IEEE, 25–32.

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

[74] Shinichi Nakagawa and Holger Schielzeth. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in ecology and evolution* 4, 2 (2013), 133–142.

[75] Anne Victoria Neale, Steven P Singleton, Mary H Dupuis, and Joseph W Hess. 1990. The use of behavioral contracting to increase exercise activity. *American Journal of Health Promotion* 4, 6 (1990), 441–447.

[76] Mark W Newman, Debra Lauterbach, Sean A Munson, Paul Resnick, and Margaret E Morris. 2011. 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*. ACM, 341–350.

[77] John C Norcross, Marci S Mrykalo, and Matthew D Blagys. 2002. Auld lang Syne: Success predictors, change processes, and self-reported outcomes of New Year's resolvers and nonresolvers. *Journal of Clinical Psychology* 58, 4 (2002), 397–405.

[78] Ann A O'Connell and D Betsy McCoach. 2008. *Multilevel modeling of educational data*. IAP.

[79] Ted O'Donoghue and Matthew Rabin. 1999. Doing it now or later. *American Economic Review* 89, 1 (1999), 103–124.

[80] Joonyoung Park, Hyunsoo Lee, Sangkeun Park, Kyong-Mee Chung, and Uichin Lee. 2021. GoldenTime: Exploring System-Driven Timeboxing and Micro-Financial Incentives for Self-Regulated Phone Use. In *In CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. ACM, New York, NY, USA.

[81] Virginia Paul-Ebhohimhen and Alison Avenell. 2008. Systematic review of the use of financial incentives in treatments for obesity and overweight. *Obesity Reviews* 9, 4 (2008), 355–367.

[82] Erin S Pearson. 2012. Goal setting as a health behavior change strategy in overweight and obese adults: a systematic literature review examining intervention components. *Patient education and counseling* 87, 1 (2012), 32–42.

[83] Aarathi Prasad, Jacob Sorber, Timothy Stablein, Denise Anthony, and David Kotz. 2012. Understanding sharing preferences and behavior for mHealth devices. In *Proceedings of the 2012 ACM workshop on Privacy in the electronic society*. 117–128.

[84] Todd Rogers, Katherine L Milkman, and Kevin G Volpp. 2014. Commitment devices: using initiatives to change behavior. *JaMa* 311, 20 (2014), 2065–2066.

[85] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2014. Personal tracking as lived informatics. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1163–1172.

[86] Heather Royer, Mark Stehr, and Justin Sydnor. 2015. Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics* 7, 3 (2015), 51–84.

[87] Alain Samson. 2017. The behavioral economics guide 2017 (with an introduction by Cass Sunstein).

[88] Manu Manthri Savani. 2018. The Effects of a Commitment Device on Health Outcomes: Reputational Commitment and Weight Loss in an Online Experiment. *International Journal of Applied Behavioral Economics (IJABE)* 7, 4 (2018), 1–20.

[89] Thomas Stachnik, Barbara Brown, William Hinds, Brian Mavis, Bertram Stoffelmayr, Dozier Thornton, and Larry Van Egeren. 1990. Goal setting, social support, and financial incentives in stress management programs: A pilot study of their impact on adherence. *American Journal of Health Promotion* 5, 1 (1990), 24–29.

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

[91] Peter M Steiner, Thomas D Cook, William R Shadish, and Margaret H Clark. 2010. The importance of covariate selection in controlling for selection bias in observational studies. *Psychological methods* 15, 3 (2010), 250.

[92] Juho Sun, Sangkeun Park, Gyuwon Jung, Yong Jeong, Uichin Lee, Kyong-Mee Chung, Changseok Lee, Heewon Kim, Suhyon Ahn, Ahsan Khandoker, and Leontios Hadjileontiadis. 2020. *BeActive: Encouraging Physical Activities with Just-in-Time Health Intervention and Micro Financial Incentives*. Association for Computing Machinery, New York, NY, USA, 17–20.

[93] Richard H Thaler and Shlomo Benartzi. 2004. Save more tomorrow^{TM}: Using behavioral economics to increase employee saving. *Journal of political Economy* 112, S1 (2004), S164–S187.

[94] Ivo Vlaev, Dominic King, Ara Darzi, and Paul Dolan. 2019. Changing health behaviors using financial incentives: a review from behavioral economics. *BMC public health* 19, 1 (2019), 1–9.

[95] Joanne Wall, Cliona Ni Mhurchu, Tony Blakely, Anthony Rodgers, and Jenny Wilton. 2006. Effectiveness of monetary incentives in modifying dietary behavior: a review of randomized, controlled trials. *Nutrition reviews* 64, 12 (2006), 518–531.

[96] Lucy Yardley, Bonnie J Spring, Heleen Riper, Leanne G Morrison, David H Crane, Kristina Curtis, Gina C Merchant, Felix Naughton, and Ann Blandford. 2016. Understanding and promoting effective engagement with digital behavior change interventions. *American journal of preventive medicine* 51, 5 (2016), 833–842.

Table 5: Multilevel regression results for featured goals with the ongoing commitment period type (Significance: $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$)

Predictor		Exercise Regularly ($R^2 = .640$) β (SE)	p		Maintain Weight ($R^2 = .358$) β (SE)	p		Quit Smoking ($R^2 = .563$) β (SE)	p		Lose Weight ($R^2 = .628$) β (SE)	p	
<Generic contract setting>													
Length	(week)	-0.002 (0.000)	< .001	***	-0.001 (0.001)	.58		0.000 (0.001)	.75		-0.001 (0.000)	< .001	***
Commitment	word count	-0.136 (0.103)	.19		0.044 (0.467)	.93		0.018 (0.116)	.88		0.054 (0.009)	< .001	***
	character count	0.031 (0.025)	.22		-0.008 (0.110)	.94		-0.004 (0.030)	.89		-0.014 (0.003)	< .001	***
	average word length	-0.176 (0.140)	.21		0.072 (1.535)	.96		0.011 (0.145)	.94		0.131 (0.033)	< .001	***
<Social commitment setting>													
Referee		0.039 (0.006)	< .001	***	0.116 (0.038)	< .01	**	0.061 (0.024)	< .05	*	0.029 (0.006)	< .001	***
Supporter		0.045 (0.008)	< .001	***	-0.040 (0.038)	.29		0.025 (0.029)	.40		0.017 (0.008)	< .05	*
<Social interaction activities >													
Photo upload		-0.005 (0.011)	.65		-0.098 (0.055)	.08		-0.082 (0.039)	< .05	*	0.013 (0.009)	.16	
Post	upload	0.119 (0.013)	< .001	***	0.121 (0.059)	< .05	*	0.122 (0.048)	< .05	*	0.091 (0.012)	< .001	***
	word count	0.002 (0.002)	.32		0.006 (0.013)	.62		0.002 (0.008)	.80		-0.004 (0.002)	< .05	*
	character count	-0.000 (0.000)	.35		-0.001 (0.004)	.86		-0.001 (0.002)	.81		0.001 (0.000)	< .05	*
	average word length	0.004 (0.003)	.18		-0.012 (0.012)	.34		-0.014 (0.011)	.22		-0.000 (0.003)	.93	
<Financial commitment setting >													
Total stake		0.000 (0.000)	< .001	***	-0.000 (0.000)	.35		0.000 (0.000)	.32		0.000 (0.000)	< .001	***
Stake per period		0.000 (0.000)	.31		0.009 (0.002)	< .001	***	0.001 (0.000)	.15		0.000 (0.000)	< .001	***
Recipient	Charity	0.531 (0.009)	< .001	***	0.229 (0.064)	< .001	***	0.439 (0.037)	< .001	***	0.487 (0.007)	< .001	***
	Anti-charity	0.685 (0.007)	< .001	***	0.342 (0.06)	< .001	***	0.583 (0.031)	< .001	***	0.626 (0.006)	< .001	***
	Friend	0.547 (0.01)	< .001	***	0.359 (0.066)	< .001	***	0.44 (0.04-)	< .001	***	0.505 (0.009)	< .001	***

A ANALYSES OF FEATURED GOALS

We additionally explored how influential variables differ across different featured goals (see Table 5 and Table 6). Featured goals are five goals given on stickK platform, with preset templates. Among 75,828 goals, we have "Exercise Regularly" ($n = 12,679$), "Lose Weight" ($n = 14,302$), "Maintain Weight" ($n = 441$), "Quit Smoking" ($n = 699$) and "Race" ($n = 878$). Table 5 lists the results of multilevel regression analyses on the four featured goals: "Exercise Regularly," "Lose Weight," "Maintain Weight," "Quit Smoking" and Table 6 lists the result of "Race." Taken our results together, we find that variables of financial commitment setting play a pivotal role in predicting success, and it is important to configure as many options (or commitment stacking) as possible. Note that the same set of independent variables from previous analyses was used except commitment description upload from generic contract setting to deal with a singularity issue.

A.1 Exercise Regularly

In "Exercise Regularly," approximately 81% of goals were non-financial ($n = 10,221$). We posit that such high percentage of goals without stake may have contributed to a low-level of average success rate (22.1%). Goals with stakes contributed an average amount of $48, and their duration was approximately 3 months. In case of referee selection, the number of self-referee was dominantly high ($n = 10,728$; 84.6%).

Our multilevel regression analyses showed that the variables of financial commitment setting (i.e., stakes per period, total stake, and stake recipient; $p < .001$), social commitment setting (i.e., friend-referee and supporter; $p < .001$), social activities (i.e., post upload; $p < .001$) and goal duration (i.e., total weeks; $p < .001$)

were all statistically significant indicators. One interesting observation is that post upload was an important contributing variable in success. Since this behavior requires constant monitoring and logging to keep track of changes, we posit that post uploads were found significant. From our analysis, it is possible to recommend good configurations for exercise-related goals; e.g., make a financial commitment that lasts up to three months, select friend-referee for accountability, invite others for cheers, and commit to social activities such as journaling the process with posts.

A.2 Maintain Weight

In "Maintain Weight," approximately 82% of goals were non-financial ($n = 361$), with a low average success rate of 22.7%. This category showed the lowest stake amount, putting $27.5 on average. As to goal duration, "Maintain Weight" also lasted about three months. Social commitment setting was lower with only 14% selected friend-referee; similarly, support selection was fairly low—a very small proportion of users ($n = 53$) selected supporters.

Our multilevel regression analyses on "Maintain Weight" showed that stakes per period ($p < .001$), stake recipient ($p < .001$), friend-referee ($p < .01$) and post upload ($p < .05$) were found significant. Unlike "Exercise Regularly," goal duration was not significant in this category. Although other goals may bring about changes for the better or worse, "Maintain Weight" aims at maintaining the same state consistently, which may require more persistence and temperance regardless of the duration. To find out the best goal configuration for maintaining weight, we extracted users only with 100% success rate. Most of these goals set financial commitment device with an average stake of $40 put to anti-charity, friend as a referee, and constant post uploads to keep track of their progress.

A.3 Quit Smoking

"Quit Smoking" showed the lowest average success rate (19.6%) with the highest average stake ($97), indicating the goal is relatively more challenging and requires a higher level of temperance for achievement. Similar to other goals, most users did not use a financial commitment device ($n = 554$; 79.3%). Average duration of the goal was less than three months. In referee and supporter selection, "Quit Smoking" also showed a very low percentage of friend-referee ($n = 116$; 16.6%) and supporter ($n = 70$; 10.1%).

Our multilevel regression analyses showed that among financial commitment settings, stake recipients were statistically significant ($p < .001$), whereas total stake amount or stakes per period were not significant. Though not strongly correlated, having a referee and uploading photos and posts was also found to be significant predictors. One interesting thing to note is that photo upload was one of the contributing variables to a success, for tracking visible changes of smoking cessation seems challenging at a first glance. Our closer inspection of photo uploads in the "Quit Smoking" category revealed that users would not necessarily upload photos related to their progress, but motivating images related to smoking cessation or even some irrelevant photos for fun. Post upload was also found to be significant, indicating that journaling process of temperance may contribute to overcoming temptations.

A.4 Lose Weight

"Lose Weight" was the most selected featured goal by users ($n = 14,302$). This goal showed an average success rate of 24.8%, while the majority of users ($n = 10,814$; 75.61%) made non-financial commitment. Users would put average $88 for a goal and the average duration was slightly over three months. Just like the other goals, most users selected self-referee ($n = 12,017$) over friend-referee ($n = 2,285$).

Our multilevel regression analyses showed that all variables of financial commitment setting (i.e., total stake, stakes per period, stake recipients) were found to have strong correlation ($p < .001$). Another strong correlation was found in variables of generic contract setting (i.e., goal duration, details of commitment description texts), friend-referee and post upload. Supporter was found to have significance as well, but to a relatively lesser extent. Surprisingly, photo upload is not a critical variable, given that tracking visible changes seems necessary in weight loss. We posit that fear of social stigma or users' perceived burden in carrying out such activities might have affected such results.

A.5 Race

"Race" refers to an individual's participation in a single event that involves race or marathon. When creating this goal, users are required to fill out information of an event name, location, event type (i.e., full marathon, half marathon, triathlon, other race), date, and target finish time (optional). 82.6% were non-financial ($n = 726$) while remaining goals with financial device put an average of $18 for an event. Some users even put $1,000 for a stake. 76.6% of goals were marked as failure ($n = 673$). As to referee selection, "Race" also had dominant selection on self-referee ($n = 772$) over friend-referee ($n = 106$). Supporter selection was also low (n = 83), accounting for only 9.45% of the goal category. Since this goal only support

Table 6: Multilevel regression model for race

Predictor		Race (R^2 = .587)		
		β (SE)	p	
<Generic contract setting>				
Length (day)		0.000 (0.000)	< .05	*
<Social commitment setting>				
Referee		0.017 (0.044)	.69	
Supporter		0.055 (0.046)	.24	
<Social interaction activities>				
Photo upload		-0.199 (0.070)	< .01	**
Post upload		0.163 (0.040)	< .001	***
<Financial commitment setting>				
Total stake		0.000 (0.000)	.25	
Recipient	Charity	0.338 (0.072)	< .001	***
	Anti-charity	0.499 (0.071)	< .001	***
	Friend	0.228 (0.060)	< .001	***

one-shot period type, we did not include the commitment period type (Is_One-Shot) effect and its interaction effects with other dependent variables. As shown in Table 6, our analyses on "Race" showed R^2 value of 0.587. Following factors of financial commitment setting (i.e., total stake, stake recipients) were found to have strong correlation ($p < .001$). Another correlation was found in goal duration ($p < .05$), photo ($p < .01$) and post upload ($p < .001$). However, including goal duration as the significant variable seems questionable, as the goal category only reflects one-day event.