Steptacular: an incentive mechanism for promoting wellness

Naini Gomes*, Deepak Merugu*, Gearoid O'Brien*, Chinmoy Mandayam*, Jia Shuo Yue*, Berk Atikoglu*
Adrian Albert*, Norihiro Fukumoto‡, Huan Liu†, Balaji Prabhakar*, Damon Wischik¶
*Stanford University, Email:{nainig,deepakm,gobrien,cmandayam,tomyue,atikoglu,adalbert,balaji}@stanford.edu
‡KDDI R&D Labs, Email:fukumoto@kddilabs.jp
†Accenture Technology Labs,. Email:huan.liu@accenture.com
¶University College London, Email:D.Wischik@cs.ucl.ac.uk

***Abstract*—This paper describes *Steptacular*, an online interactive incentive system for encouraging people to walk more. A trial offering Steptacular to the employees of Accenture-USA was conducted over a 6 month period. Over 5,000 employees registered for the program and close to 3,000 participants wore USB-enabled pedometers; from time to time they plugged their pedometer into a computer to upload hourly step counts to a website; and the website had a range of features to encourage more walking. These features included monetary rewards which were randomly redeemable through a simple game, and a social component. We describe the system and present preliminary findings about the effectiveness of each of these features in encouraging physical activity.**

I. Introduction

A recent UN conference addressed the problems of non-communicable diseases (mainly cardiovascular diseases, diabetes, cancer, and chronic respiratory diseases) [1]. It estimated that the worldwide death toll from non-communicable diseases is 36 million premature deaths per year, more than 60% of the total death toll. Further, 80% of the deaths from non-communicable diseases are in developing countries [2]. It found that these are predominantly 'lifestyle' diseases, best dealt with by promoting physical activity and better diet rather than by medical treatment; i.e., prevention rather than cure. The cost of non-communicable diseases world-wide is estimated to reach $47 trillion by 2030 [3].

There have been several responses to this crisis, ranging from city-wide and national wellness challenges [4], [5] to corporate and institutional wellness programs [6]. In [7] Steve Burd, CEO of Safeway, lays out the value of a Safeway wellness program, evaluated during 2005–2009, in terms of employee health and medical insurance costs. Burd credits this program with keeping Safeway's healthcare costs nearly constant over a four year period when that of other companies rose by 38% during the same time. A long-term study [8] of a wellness program offered by Johnson & Johnson in 1995 estimates that about $224.66 per employee per year could be cut from medical care expenditures, while Safeway's program reduced annual premiums by $780 per employee.

We built *Steptacular*, an online interactive system for promoting walking. We ran a 6 month trial as part of Accenture's employee wellness scheme, with over 5,000 employees registered and close to 3,000 pedometers coming online. The scheme was open to most of the employees of Accenture–USA and their spouses. We paid out a total of $238,000 in rewards for walking, equivalent to $159 per person per year. (In the U.S., the amount of money given to physical activity in wellness programs ranges between $50 to $200.) The net effect was a high enrollment, a frequent engagement with the system, and an improvement in the average steps per user per day.

The obvious hypothesis is the more people are paid to keep active, the more they will keep active. We found, however, that when it comes to persuading people to change their behavior, it's not just the amount of money you spend, it's how you spend it. There are three aspects in which network technologies can help get the biggest impact:

- an accurate and simple data collection mechanism[1]
- interactivity—fast feedback and immediate rewards for good behaviour
- a social component—listing a user's peers, and showing their activity and their winnings, can have a big impact on the user's effort and commitment

In Section II we explain how Steptacular works, and how it incorporates these three aspects. In Section II-C we describe our experience of deploying Steptacular, and present some summary statistics. In Section III we analyse the effectiveness of different aspects of Steptacular, by a series of before/after comparisons: changing the amount of incentive payment, introducing a social component, and introducing a narrative element. We also describe another incentive scheme run at Accenture, the *10k Challenge* and compare Steptacular with it.

We conclude in Section IV with a discussion of how this work relates to ongoing work on 'gamification', and describe our plans for future work.

Related work

Steptacular has three features: a focus on technology, a large number of participants, and monetary incentives. Wellness programs such as the 10k Challenge [5] and Safeway's program [7] have been administered at scale, but they have either relied

[1]The pedometer we used in Steptacular was not easy enough to wear and use for many participants. In future, we believe mobile phones can play a key role in data collection.

on self-reporting [5] or have not made extensive use of network technology [7]. On the other hand, some pedometer makers [9] make strong use of technology by using accurate sensors and providing users with accounts for tracking their activity. However, they don't directly administer a wellness program and give monetary incentives. So our work complements existing programs in some interesting ways.

II. The Steptacular Scheme

A. The user experience

Users registered online at steptacular.org. Registration was limited to Accenture-USA employees and their spouses or domestic partners; this was enforced by email verification. A registered user purchased an Omron HJ720IT pedometer for about $33.[2] The user also installed PC software, which retrieves pedometer readings over USB (recorded hourly for up to 42 days) and uploads them to steptacular.org.

Walk, earn, redeem. Users were encouraged to wear the pedometer and walk. All the steps they took—to the parking lot, to the conference room, etc—were counted up to a maximum of 10,000 steps per day (approximately 5 miles). Step counts were converted into *credits* through a formula that increased super-linearly with the number of steps. Users were also assigned a *status level* each week, silver, gold or platinum, based on the number of steps walked in the previous week. Users could then redeem their credits for money, using the incentive interface webpage described below. In addition to this webpage, the website consisted of three other parts: the main user account page, 'my steps', and an informational section. Figure 1 shows the system block diagram.

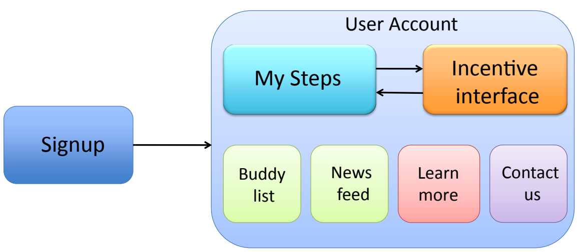

Fig. 1. System Block Diagram

Main user account page (Fig.2). A user who logs in to Steptacular starts on this page. It displays a summary of the total steps taken by the user to date, the total rewards earned, and the current status. It also shows:
1. The user's friend list. Users were able to invite other users to become friends; when the invitation was accepted both users were added to each other's friend list. The friend list shows each friend's current status, number of steps in the previous week, and total rewards won.
2. The newsfeed. This is a constantly updated and timestamped newsfeed alerting users about uploads ('Someone uploaded X steps just now') and rewards won ('Someone just won $5 on the Gold board'). When the someone is a friend, they are named, otherwise the alert is anonymous.

[2]The first 4,000 registered participants were given a $10 Amazon gift coupon to buy the pedometer. About 2,000 of these coupons were used to buy the pedometer.

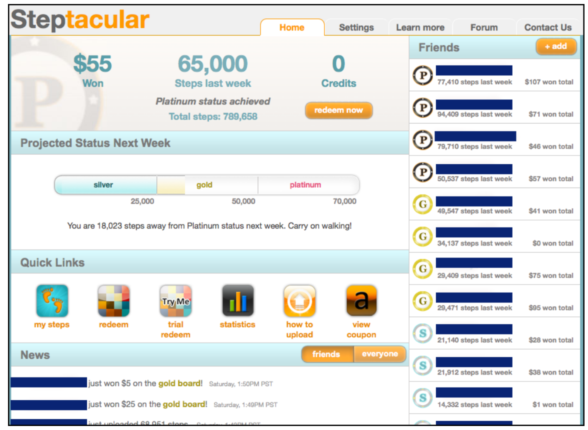

Fig. 2. User Account: Landing Page (names redacted)

My Steps page (Fig.3). This shows the daily step count of the user over time, the calorie expenditure from these steps (estimated using a simple formula based on step count, height and weight), the average number of steps taken by all users each day, and a history of credit redemption.

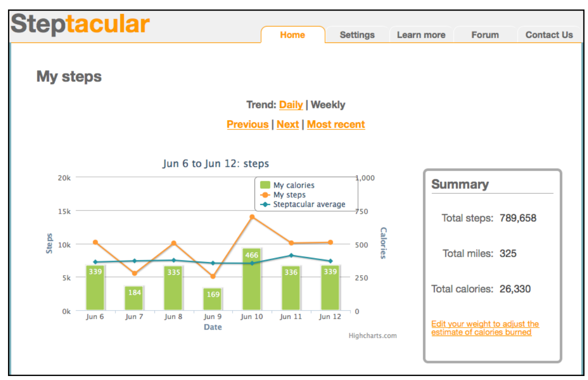

Fig. 3. My Steps page

Incentive interface page (Fig.4). Users could choose to redeem their credits in two ways—deterministically or randomly. The deterministic option gives them $15 for every 5000 credits. The random option offers rewards ranging from $1 to $100 via a chutes and ladders boardgame, in which die-rolls cost credits and payouts are associated with certain tiles. From a mathematical perspective, the game is a Markov chain which advances when the user rolls the die. From a financial perspective, the game is a method for implementing a raffles- or sweepstakes-like scheme in a distributed fashion. From the user perspective, is a fun way of making random rewards both intuitive and transparent.

There are three gameboards—silver, gold and platinum—with increasing rollcosts and increasing average dollars per credit spent. A silver-status user can only play on the silver board; a gold-status can play on either silver or gold; and a platinum-status user can play on all three.

The gameboards have three different types of reward tile: extra credits, money, or boost. The boost tiles double the credits earned per step, for steps walked over the next four days. Some of the credit and money tiles offer straight rewards, e.g. $5. Others are two-value tiles, e.g. $1/$5. When the user's gamepiece lands on one of these, the system chooses either the low value or the high value, and the user does not know in advance which will be chosen. The system makes its choice based on the amount of money it has in reserve. This allows it to limit its total liability; e.g., in the event of lots of users by chance landing on reward tiles, or lots of users walking more steps than usual. At a higher level, it is a method for ensuring a zero-sum payout mechanism across the participants when each one of them redeems their credits in parallel play on the boards.

Users have a choice of how to advance on the gameboard. They can roll a regular die. Or they can choose the flip die, which is a tray of 6 tiles containing the numbers 1–6 in random order which the user can flip open one at a time. Or they can play some or all of their credits in one go by selecting Autoplay.

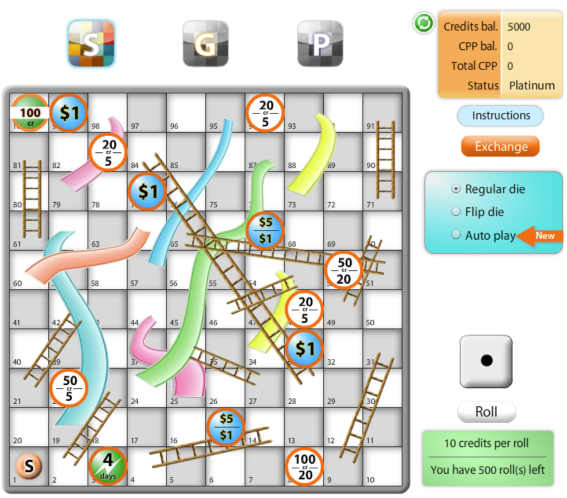

Fig. 4. Chutes and ladders game

Informational pages. Several informational and help pages were provided, such as Learn More, Overall Statistics, a forum, and a FAQ. Moreover, participants could contact the Steptacular team via email and get a response in 2 business days. Over 3,000 emails were received during the pilot.

B. Implementation

The system was hosted on Amazon's Elastic Compute Cloud (EC2) as an extra large instance. It comprised of a MySQL database to store user data and record all transactions, an Apache webserver to host the web portal and a Java server to handle redemption. The chutes and ladders game was available on the web portal as a Flash application. A user's data was fetched from the database to the web server using PHP.

C. Deployment

The pilot was launched on the 22nd of March, 2011 to about 100 members of Accenture Technology Labs for beta testing. The final version of the pilot was launched on the 29th of March, 2011 to employees of Accenture, USA (about 30,000 employees).

Not all the features described in Sec.II-A were present at launch. Fig.5 shows when each feature was introduced, and the overall timeline of deployment. In addition to the features described above, we introduced a 'moon shot' on 10 May. This was a graphic representation of total steps taken by all users, depicting this total distance by means of a rocket on its way from the earth to the moon, and shown on the front page of the website. The contrails of the rocket showed the number of steps contributed by each participating offices. Moon landing was on 19 May, whereupon the front page reverted to its old form.

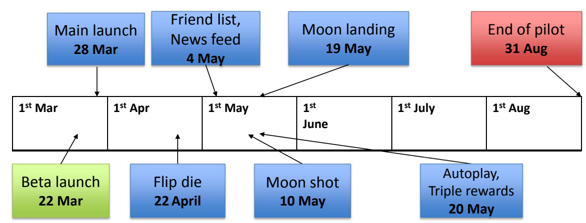

Fig. 5. Timeline

Over the course of the pilot, 5,093 people signed up for the program. Of these, 2980 had uploaded their steps at least once. In all, 1,841,983,929 steps were entered into the program. The overall daily average step count improved from 6,584 in early April to 7,822 by the end of the program (though this figure is inflated by the fact that lower-stepcount users were more likely to drop out; when we adjust for this the final average step count is 7,087.) Most users tend to carry their pedometers with them on weekdays, and many tend not to use them on weekends. However, Saturday had the highest average stepcount, at around 7678 steps over the duration of the program.

At the end of the pilot, a 15-question survey was sent out to the participants. It elicited 722 responses. One of the questions asked users which features they found compelling: 79% cited the monetary reward, 64% cited recording and tracking progress, and 45% cited 'gadget+online = fun'. Over 70% said they walked more during the pilot than before. When compared with the 10k Challenge program previously offered by Accenture,[3] about 55% said they thought Steptacular was more generous and 15% said it was less generous. Further, 64% said they would stick with Steptacular for a longer period of time than with the 10k Challenge. Of the people who

[3]More on this comparison in the next section.

stopped using Steptacular before it ended (and 164 answered this question), losing the pedometer and too little reward money were stated as the primary reasons followed closely by a dislike for the pedometer.

III. Analysis

In this section we analyse how networking and communications technologies can enhance an incentivization program. It is not surprising that paying money for walking leads to more walking—but can we quantify the additional benefits of easy data collection, game-like interactivity, and social networking?

The classic way to address this question would be to conduct a series of randomized controlled trials. This was not practical for us because our approach sought to build a community of users, consisting of friends and colleagues, which makes it difficult to construct randomized trials without creating a sense of unfairness. However, we are still able conduct three types of analysis: (1) simple analysis of correlations between user characteristics and behavior, (2) before-and-after analysis of the effect of adding or changing a feature, and (3) using the inherent randomness of the redemption game to analyse the data as a collection of controlled randomized 'mini-trials'.

In the rest of this section, we report on the online social incentive effect, the overall effectiveness of easy data collection and game-like interactivity, and the relationship between money and gameplaying.

A. Social incentives

We introduced the friend list and newsfeed features, described in Section II, 37 days after launch. By the end of the trial, 1186 users had added friends, and among these the median number of friends was 2 and the mean was 2.65, giving a total of 1569 friendships. 50% of friendships had been added within 19 days after the feature was launched, and 75% within 47 days.

We can distinguish two social effects. First, by looking at the week-by-week performance of the group of users who eventually added friends, and comparing to the group of users who did not—but only looking at their performance data before they had added any friends—we can find the correlation between performance and having friends participating in the program. Call this the *social effect*. Second, by comparing the two groups but only looking at the data after the first group had added friends, and subtracting the social effect, we can estimate the impact of providing an online friend system. Call this the *online social effect*. The online social effect reflects the added benefit of e.g. being able to easily see exact details of your friend's performance, rather than just discussing it in vague terms around the water cooler. This desire for seeing an exact comparison is why we prefer to see runners all run a race together, rather than record their times sequentially.

(In many systems it will be hard to distinguish between the social and the online social effect. For example, in geographically dispersed systems, it may be that the only way to build up social communities is via an online system.)

The social effect is highly significant on all the measures we looked at: steps per day, user retention, and user engagement; though beyond 6 friends, further friends have no impact. The online social effect is 34% of the social effect on steps per day, i.e. by providing an online tool to reflect online friendships we boost the social effect by 34%. We don't have appropriate data for analysing the online social effect on the other two measures. We now give detailed analyses.

a) Effect on steps/day: The social effect on daily step count is an extra 781 steps/day (std. error 104), and it is broken down as follows:

num.friends	1	2	3–4	5–6	7+
extra steps/day	525	664	1045	1561	1531
std.err	132	206	214	300	325

The effect of having a spouse in the system is an extra 244 steps/day (std. error 151), which is not significant. The online social effect is an extra 266 steps/day (std. error 32). The statistical tool beind this analysis is a linear mixed effects model, with a random per-user intercept.

b) Effect on user retention: We now analyse how likely it is that a user stays with the program. Define the 'survival time' for a user to be the time from the first walkday to the last walkday, where a walkday is a day on which the user wears the pedometer and some time later uploads the stepcount. If the last walkday is in the final week of the program, treat that user's survival time as censored. The social effect can be measured by the probability of surviving for some given number of days, as a function of the number of friends. Obviously the two should be correlated, since the longer you are in the system the more time you have to make friends; in order to get a true reading of the social effect we have therefore restricted the analysis to friendships made in the 21 days after the social feature was launched, and to users who stuck with the program for at least those 21 days. Among these users, the probability of surviving for at least 150 days is

num.friends	0	1	2	3–4	5–6	7+
prob.	58%	69%	63%	81%	78%	91%
std.err	1.5%	2.4%	5.2%	3.0%	6.8%	3.4%

The online social effect on retention cannot be measured from the data we have—by definition, users who later made friends did not drop out before they made friends.

c) Effect on user engagement: We now analyse how likely it is that a user logs into the website each day. The probability of logging in to play or upload on a given day was 28% at the beginning of the trial and it decreased to around 15% over the course of the trial; however the social effect was fairly stable from week to week and it was

num.friends	1	2	3–4	5–6	7+
extra prob.	0.9%	-0.8%	5.2%	8.1%	9.6%

The 1- and 2-friend effects are not significant, but the others are (in a linear mixed-effects model). The online social effect is not significant.

We started full recording of all logins (including logins just to view your friends' status) on day 66 of the trial. The total login probability is 2.6% higher for users with 1 friend, 2.2% with 2, 7.3% with 3–4, 10.7% with 5–6, 10.6% with 7+. Because we do not have full login records prior to the introduction of the friend list feature, we cannot estimate the online social effect. We expect it to be significant—users presumably add friends for the purpose of viewing their friends' status (and to boast).

B. Monetary/game incentives

Users found Steptacular to be more engaging than the 10k Challenge incentive scheme that ran previously, as discussed in Section III-C below. This is likely to be due, at least in part, to the interactive 'gamified' reward scheme we implemented. This leads to the question: how does gamification impinge on the incentive effect of money?

The structure of the trial gives us three ways to look at this. (a) Part way through the trial, we increased the payouts. What was the effect? (b) Some players were lucky and landed on boost tiles, which doubled their expected payout rate over the next four days. Did they respond differently to players who missed out on the boosts? (c) The chutes and ladders game is random, which means that for the same amount of effort some players were lucky and won more than others. Did they respond differently?

The detailed results are described below. Taken together, they show that the more reward money users perceive, the more effort they put in. The operative word here is *perceive*: humans are notoriously bad at estimating probabilities, especially of rare events [10]. When we changed the boards to double the payout rate, most users did not realize, and we saw virtually no impact on performance. Nor did luck of the die have any significant impact. On the other hand, when users were explicitly told via in-game boost tiles that they would be earning double points for a period, performance increased significantly for that period.

It might be suggested that users were responding to the intrinsic 'gamified' fun of the boost, and didn't care about extrinsic extra payouts. This seems unlikely, for two reasons. First, survey respondents certainly said they valued the payouts: 79% said that money was one of the most important features encouraging them to participate, compared to 29% who said it was the fun of the game. Second, when asked how they would respond if the payouts were doubled, 73% said they would walk more.

a) Increased payouts: On day 54 of the trial, we more than doubled the payout rates. The empirical average payout rates (¢ won, divided by credits played minus credits won) were

	silver	gold	platinum	overall
days 0–53	0.38¢	0.42¢	0.53¢	0.43¢
days 54–156	0.69¢	0.78¢	1.11¢	0.89¢

Average steps/day increased somewhat, from 6640 over days 0–53 to 6839 over days 54–156. In order to decide whether this is significant, and in order to distinguish the effect of changing the payout rates from the overall effect of 'getting used to the scheme', we can look at the variation in steps/day over different subperiods:

days	0–53	54–79	80–105	106–131	132–156
steps/day	6640	6773	6928	6754	6934

We can conclude from this that the increased payouts had an insignificant effect.

We believe users were aware of the change. They were informed by email that rewards would triple, and told that for the sake of fairness their existing point balance would be divided by 3. There was a big spike in playing in the week before the change, after the email went out. Of the 702 users who responded to the survey, 65% said they were aware of a change.

However, the survey results also show that most users didn't actually perceive how much we increased payouts: 42% of survey respondents said that they noticed an increase, and only 6.8% said that payouts more than doubled. This 42% did increase their steps/day, but no more than the population as a whole. With the benefit of hindsight, we would choose to increase the value of the payoffs—say, by doubling some of them—rather than merely increasing number of tiles that paid off. This would have been more noticeable by the participants.

b) Boosts: Boost tiles, described in Section II, give four days of double credits to users who land on them. We can treat boosting as a classic randomized trial: among the cohort of users who land on a given boost tile, some get the boost and others do not, and the allocation is governed by the roll of the die.

A boosted user walks 586 steps more than average (std. error 147) on the day after winning the boost, and the boost effect declines over the four days. On day 5, once the boost is over, steps/day are no different to before the boost. A boosted user is more likely to upload steps and to play while boosted, roughly 6% greater chance for both activities, for days 1 to 5. By day 6, the probability reverts to its pre-boost level. (These statistical analyses use linear mixed effects models.)

c) Luck: Some users are lucky with their die rolls and earn lots of money. We now analyse whether luck has an impact on future performance. We measured luck by total reward money won per credit played, over the month of June, days 65–94, and restricted attention to users who played at least 2000 credits. (A long window is needed in order to get a measure of luck which isn't all-or-nothing.) We measured performance by change in steps/day from June to July, and restricted attention to users who walked 7000–8000 steps and $\geq$20 days in June, and $\geq$15 days in July. (This restriction is so that we are comparing like users.) This left 172 users. Among these, there was no significant correlation between luck and improvement in performance.

C. Steptacular v. 10k Challenge

Prior to the launch of Steptacular, Accenture employees had access to another wellness incentive scheme, the 10k Chal-

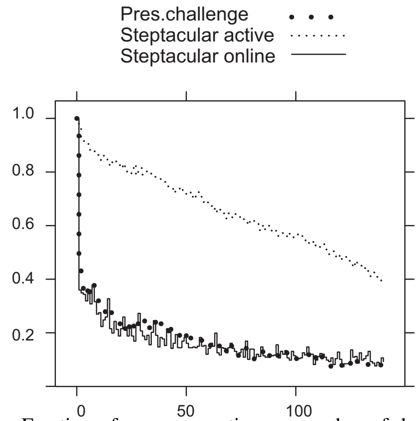

Fig. 6. Fraction of users are active vs. number of days in program

lenge. Users visit www.presidentschallenge.org and manually enter the duration, intensity and type of each physical activity they do, and are awarded points. If they achieve a target number of points in a year, Accenture pays them $100.

There were 390 users common to both schemes. From these users, we have a controlled look at how the design of an online wellness experience affects participation. Figure 6 shows what fraction of these common users participated, on each day after they started: did they log an activity for that day on the 10k Challenge website, did they wear their pedometer for Steptacular, did they upload steps or play the redemption game? The plot shows that the frequency of online activity is remarkably similar for the two schemes, whereas the frequency of recorded physical activity is much higher for Steptacular. In other words, ease of data collection is a crucial component of networked wellness systems, and fun and games on a website are not enough on their own to increase participation.

Among the 390 common users, 53% had discontinued the 10k Challenge within 100 days of starting, and 36% had discontinued Steptacular. Section III-A suggests that the social features contributed to Steptacular's greater 'stickiness', but it may also have come from the fun of the redemption game or the fun of seeing and tracking your data. Indeed, over about 73% of the credits were redeemed through die rolls, while 25% were auto-played and slightly over 2% were redeemed deterministically. With easier data collection, stickiness would have been even higher: among survey respondents who gave a reason for discontinuing, 58% complained either about the bother of wearing it or the bother of replacing a lost pedometer. 73% of survey respondents said they would stay with Steptacular longer than the 10k Challenge. Comparing the user perception about payouts, 55% said Steptacular was more generous, 25% said Steptacular was more generous but they prefer a non-random payout, 5% said the two schemes were similar, 15% said Steptacular was less generous.

IV. Conclusion

The paper described a wellness pilot conducted for the employees of Accenture-USA. The main features of the pilot were: (a) an informational component consisting of a precise measurement of a participants's effort, (b) a social component consisting of a participant's friends' performance, and (c) a monetary component.

We found that the monetary and social components had a larger effect on a participant's performance and 'stickiness' with the program. While participants valued extrinsic rewards in the form of the money they won, the social context within which they participated in the program provided a significant motivation and increased their commitment. Money was paid out through an online game; gamification introduced an element of fun, but it also confused users about how much money they could actually win.

Based on this study, we believe that a well-designed incentive scheme ought to: employ technology that is very easy to use, provide clear data on a participant's effort, provide an opportunity to redeem meaningful sums of money for the effort they expend, and create a social structure for participants from which they can derive motivation and increase their commitment. We believe that elements of our work can be used to develop interesting wellness incentive mechanisms in technologically less sophisticated contexts—e.g., in the developing world. Indeed, the prevalence of cell phones makes it possible to devise interesting and engaging programs.

Acknowledgment. We thank Wynn Pott and Julie Wilkes of Accenture's Live Well program for their generous support of this pilot, and for their insights and feedback during all phases of Steptacular.

References

[1] "UN conference on non-communicable diseases," Sep 2011. [Online]. Available: http://www.un.org/apps/news/story.asp?NewsID=39642
[2] "WHO fact sheet on non-communicable diseases," Sep 2011. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs355/en/index.html
[3] "Non-communicable diseases to cost $47 trillion by 2030," Sep 2011. [Online]. Available: http://www.weforum.org/news/non-communicable-diseases-cost-47-trillion-2030-new-study-released-today
[4] "Chicago's "wellness" program for city employees," Sep 27 2011. [Online]. Available: http://www.anh-usa.org/chicago-wellness-program-imagine-if-this-went-national/
[5] "The president's challenge," Accessed Oct 10th 2011. [Online]. Available: http://www.presidentschallenge.org/celebrate/active-lifestyle.shtml
[6] R. Goetzel and R. Ozminkowski, "The Health and Cost Benefits of Work Site Health-Promotion Programs," *Journal of Public Health*, vol. 29, 2008.
[7] "How safeway is cutting health-care costs," June 2009. [Online]. Available: http://online.wsj.com/article/SB124476804026308603.html
[8] R. Ozminkowski, D. Ling, R. Goetzel, J. Bruno, K. Rutter, F. Isaac, and S. Wang, "Long-term Impact of Johnson & Johnson's Health & Wellness Program on Health Care Utilization Expenditures," *Journal of Occupational and Environmental Medicine*, vol. 44, no. 1, 2002.
[9] "Fitbit," Accessed Oct 10th 2009. [Online]. Available: http://www.fitbit.com/
[10] D. Kahaneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, no. 2, 1979.