GambleFit: Improving the Efficacy of Wearable Fitness Trackers Through Gambling

Nikhil Charles, Devon D'Apuzzo, Fahad Khan, Rovik Robert

Abstract— **In recent years, the popularity of wearable fitness trackers has exploded. While their accuracy has been the subject of many studies, their efficacy as tools for improving weight loss or physical activity has not received as much attention. Some recent papers have touched on the subject, however, indicating that fitness trackers alone are not useful for maintaining recommended activity levels in the long term. In this paper we present a fitness tracker interface, GambleFit, that employs a betting and guessing mechanism grounded in the theory of planned behavior to better encourage behavior change. This is accomplished by teaching users how to set realistic fitness goals and rewarding them with virtual currency. Due to limited access to equipment, we had a very small sample size of users. Because of this we were unable to draw conculsive results about MVPA per user.**

I. INTRODUCTION

With recent advancements in mobile computing, software has become more efficient, hardware has become both smaller and cheaper, and a broad range of devices have become "smart". After the 2007 release of the iPhone, smart devices have come a long way, integrating more computational ability, encompassing a larger span of devices, and becoming more ubiquitous in our daily lives. Fortunately, for health professionals and programmers alike, these smart devices also ship with a large variety of sensors. Using data analytics on sensor data obtained through passive sensing, these smart devices can be used to profile user's activity levels and fitness state. Fitbit, Google Fit, and the Apple Health app have recently adopted this method of tracking fitness levels for users without requiring excessive self reporting. Simultaneously, the growth of smart wearable device - namely smart watches - has only added to the available sensor suite and data. While the accuracy of the data collected by these trackers has yet to be fully understood, devices such as the Fitbit can use their sensor suite to relay step counts, calories burned, and hours of sleep. This information is relayed to the user in the hopes that it will help them understand their current level of fitness. The homepage of the Fitbit website states that "Fitbit motivates you to reach your health and fitness goals by tracking your activity, exercise, sleep, weight and more." [7] This sentiment is generally echoed for most activity trackers.

However, recent studies suggest that physical activity monitoring devices such as a Fitbit are not completely effective in promoting health and fitness. When added to simple behavioral intervention, such as a fitness plan, some studies certainly show Fitbit-like fitness trackers aiding in the process of weight loss. However, these studies are mostly confined to the short-term and the benefits are not prevalent in longer studies [8]. The reasons for this are plentiful, but mostly stem from behavioral theory. Namely, the theory of planned behavior and social cognitive theory suggest that the reason fitness trackers fail is because they do not aid in long-term habit formation [10]. Habit formation requires a user to believe that they have control over their ability to hit fitness and weight loss targets. Attaining this level of control requires intentional thinking and participatory learning of how to set attainable goals. Users must be able to internalize adequate goals and truly understand their progress for behavior to change occur over a longer term [4]. Both behavioral theories also highlight the importance of short term benefit over long term gains. A user is a lot more likely to undergo behavior change if there is a short term benefit to performing the action. A long term benefit, such as the promise of being more fit in the future is generally not enough to encourage behavior change, hinting at why fitness trackers may not be effective at promoting behavior change [10].

Using these theories, it was clear that a key problem with fitness trackers is the absence of adequate goal setting. Fitbit, Google Fit, and Apple Health all require the user to set a fixed goal when they first start the application. This goal is set at a time when the user in not aware of what an adequate fitness goal consists of and is rarely updated. Goal setting in fitness regiments has been shown to be a major component in a successful program. [3] Thus, in this paper, we introduce the concept of gamification to fitness trackers. The product, GambleFit, uses a betting mechanism to engage users in the process of learning how to set attainable goals. A virtual currency, Acticoins, is used to bet on fitness goals. We hope that the reward of gaining Acticoins for setting attainable goals allows users to believe that they have control over their fitness plan. A daily fitness guessing mechanism is also implemented to further teach users to accurately gauge activity levels and by extension be more intentional in considering their daily activity plans in relation to their goals. Acticoins are also used as a short term reward for meeting daily goals in an attempt to provide a more short-term benefit to participants. We test the GambleFit system against an application that simply displays static fitness information as most fitness trackers do. For our study, we use minutes of moderate to vigorous physical activity (MVPA) as our metric of activity level.

II. RELATED WORK

There are numerous ways in which people use gaming in mHealth. Douris, Peter C., et al. evaluates the physiological and psychological responses of college-age adults playing the Nintendo Wii and finds that the Wii Fit game is not as effective as traditional exercise for improving positive well-being [6]. A similar study used the Freedroid RPG game on Android to ”derive behavioral patterns and context from human motion,” which is useful for detecting periods of physical activity [2]. GambleFit differs from these approaches by using gaming to increase mindfulness of users in relation to their physical activities with the goal of increasing their engagement with the system.

Not many studies exist on the impact of game techniques in improving physical activity. Tabak et. al. explores the claims that gaming applications improve physical activities but ends up inconclusive on the effectiveness. While finding a repeated occurrence of goal-setting as a means of motivation and rewards and competition as game elements, clinical evidence on their effectiveness was limited with only two randomized controlled studies found. Furthermore, a lack of theoretical basis for the game frameworks was found amongst the applications studied, suggesting that they do not directly address behavioral health theories [12].

Direito et. al. explored a similar space, calling them Apps for IMproving FITness (AIMFIT) but categorized the applications into immersive and non-immersive applications. Conducted over 8-weeks, this study compared primarily ”cardiorespiratory fitness, objectively assessed as time to complete the 1-mile run/walk test at 8 weeks. Secondary outcomes were physical activity levels (accelerometry and self-reported), enjoyment, psychological need satisfaction, self-efficacy, and acceptability and usability of the apps”. The study found no significant intervention effects as a result of the games, and cited that the ”more tailored approaches, which are dynamic and responsive to changes in physical activity over time” would be useful for further studies [5].

Both of these studies encourage us to firstly revisit fundamental health behavioral theories to form the basis of our own game, as well as incorporate elements that are not yet studied by the wider research community into the game for the purpose of studying their effectiveness. In this regard, we consider gambling as a new feature and base it on the transtheoretical model as well as social cognitive theory.

In the transtheoretical model, short term motivation is important in the action stage and learning how to set goals is important in the maintenance stage [10]. Bandura’s Social Cognitive Theory also place a strong emphasis on improving self-efficacy as well as modifying social environments to improve behavior. These are the theoretical frameworks we use to guide the elements we incorporate within the larger gambling mechanism [4].

III. SYSTEM

The GambleFit platform is comprised of three main components: the Android Mobile application, the Android Wear application, and the back-end server application. The mobile application serves an an interface for the user and is the part of GambleFit with which the user interacts. The Android Wear application runs on a smartwatch and collects the user’s accelerometer and Google Fit data for the day. This data is formatted and transferred via HTTP requests from the mobile application to a back-end mySQL database. Then the MVPA can be requested. The back-end server calculates the MVPA for each user each day. Finally, the server returns the MVPA calculation to the application to show the user their MVPA and the interface calculates and displays their currency.

A. Android Mobile Interface

A user interacts with GambleFit through the Android Mobile application.The structure of this game is illustrated in figure 1.

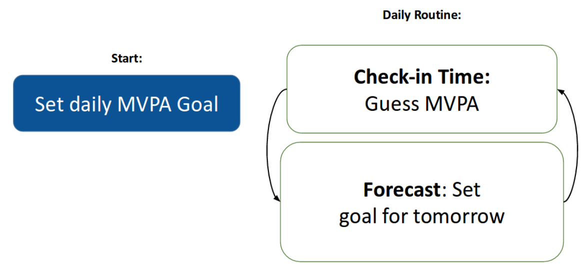

Fig. 1. Basic structure of GambleFit mobile interface

When running the application for the first time, a user is prompted with a form asking them how many minutes of moderate to vigorous physical activity (MVPA) they want to engage in per week. This form is only presented to the user the first time they run the application.

After the initial setup, the user is then asked to check-in with GambleFit every night. At the nightly check-in time, the user must guess how many minutes of MVPA they engaged in that day. After guessing, GambleFit shows the user how close they were to achieving their goal. If they guess correctly, they earn Acti-coins, the virtual currency used in the game. If not, the user loses a fraction of their Acti-coins. This interface is seen in figure 2.

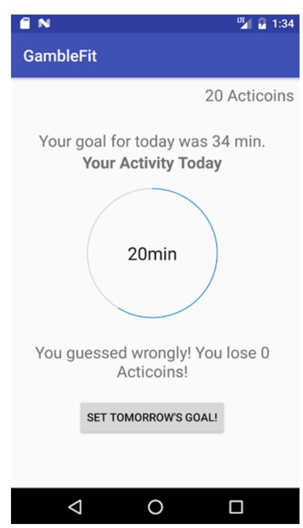

Fig. 2. GambleFit verifies your guess

Finally, the user is prompted to set a new goal for the following day. If the user desires, they can gamble some of their Acti-coins to try and earn extra Acti-coins. This is seen in figure 3.

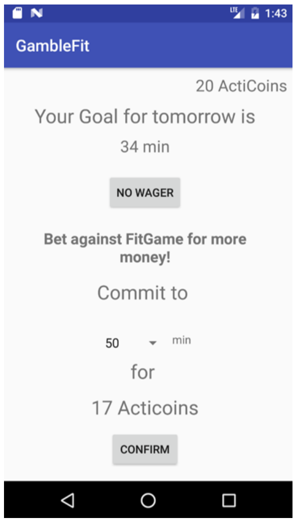

Fig. 3. GambleFit wagers against you

B. Android Wear Application

GambleFit collects information about a users physical activity through the use of an Android smartwatch. The smartwatch logs activity data through Google Fit, which contains data like the user's step-count and heart rate. The smartwatch also measures raw accelerometer data from the watch itself and transfers this data to the mobile application at regular intervals.

C. Backend Server

At regular intervals during the day, the user's Android phone sends step count, heart rate, and calorie data to the Google Fitness Store, data which is automatically collected using sensors on the smartphone and the smartwatch. This setup is illustrated in figure 4. When it is time for the user to check in to GambleFit, GambleFit makes data requests to the Google Fitness Store to obtain the user's activity data for the day.

At the end of each day, the mobile application transfers the data collected from the Google Fitness Store to the backend server over a HTTP request, where it is stored in a MySQL database. The back-end server then uses this data to calculate the user's daily minutes of MVPA (mMVPA), which is described in the next section. Figure 5 shows how Google Fit interacts with the other elements in our system.

D. Feature Extraction and Segmentation

Using Google Fit, we are able to obtain the following information about a user's activity:

- Number of Steps
- Maximum Heart Rate
- Minimum Heart Rate

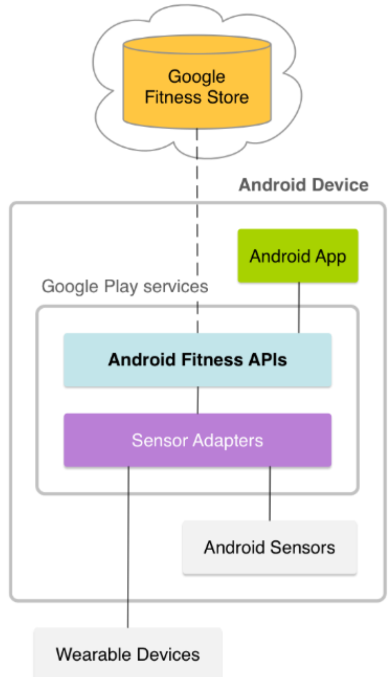

Fig. 4. Basic structure of Google Fit

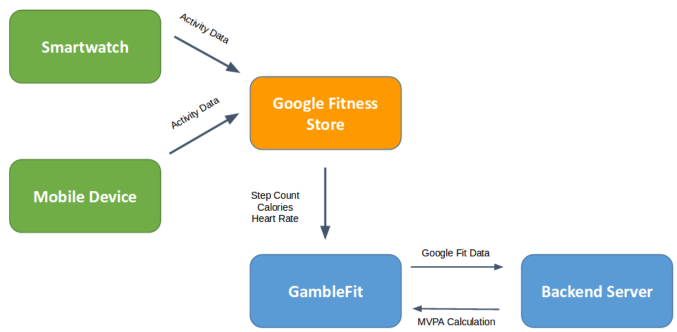

Fig. 5. Interaction between devices, Google Fit, back-end server, and GambleFit

- Average Heart Rate
- Calories Burned

Initially, we had also intended to use accelerometer data collected from a smartwatch to calculate a composite acceleration value with an algorithm originally proposed in [13]. This algorithm is shown in equation (1), where Q and P are defined in equations 2 and 3, respectively.

$$K_m = \sqrt{\frac{1}{n-1}[Q - \frac{1}{n}(P)]} \tag{1}$$

$$Q = \sum_{i=0}^{n} x_i^2 + \sum_{i=0}^{n} y_i^2 + \sum_{i=0}^{n} z_i^2 \tag{2}$$

$$P = (\sum_{i=0}^{n} x_i)^2 + (\sum_{i=0}^{n} y_i)^2 + (\sum_{i=0}^{n} z_i)^2 \tag{3}$$

While testing this equation in our application, we found that collecting accelerometer data demanded excessive amounts of power from the smartwatch and therefore was not suitable for our task. Additionally, we found that the accelerometer data is highly correlated with step count data calculated by Google Fit. Consequently, we opted to disregard accelerometer data in favor of step count for our final prototype.

E. MVPA Calculation

Since heart rate, steps, and calories are all reported in intervals, we chose to calculate MVPA in intervals as well. We chose to make our time windows 5 minutes since the best time resolution we could get from Google Fit was 5 minutes. We wanted the windows to be as small as possible to allow for the greatest degree of precision, but big enough to accurately handle the uncertainty of the Google Fit data.

The intervals that Google Fit reported were not consistently 5 minutes, nor were they consistently spaced, so the first task of the MVPA Calculation was to bucket each raw data point to our discrete scale. Whenever a user requested their MVPA, the back-end created a list of 'bucket' objects. These had a start and end time, along with fields for each of our features. Each bucket was 5 minutes in length. By default the first bucket was from 5 AM - 5:05 AM and the last bucket from 9:55 PM - 10:00 PM. The program iterated over the raw data intervals and had to decide how to properly bucket each observation. When it encountered calories and steps, it first checked how many buckets the interval overlapped with. If it only overlapped with one, it placed the entire value in that bucket. If it overlapped with more than one, it proportionally divided up the raw amount to each bucket it overlapped with.

Heart Rate measurements from Google Fit were treated differently. It first checked if there was already a minimum, maximum, or average value reported for the buckets the raw observation overlapped with. If there wasn't any value, it simply made the bucket's value the observation value. If there was, it compared to the observation value. For max, if the observation value was bigger, it replaced the current value. Likewise, for min, if the observation value was smaller, it replaced the current value. Finally, for average, it averaged between the current value and the observation value.

Once all of the raw data is bucketed, there are usually a large number of missing values. For step count, we simply made missing values 0, since Google Fit does not report steps if there were no steps. For calories, we could estimate Google Fit's reported value for resting calories. We would place this value in for any missing calories value. Heart Rate had the highest number of missing values. Originally we chose to propagate the most recent reading through until we reached another reading, but that didn't work well. We had a few cases where a high heart rate was reported followed by no readings for a while. This high heart rate shouldn't be propagated through indefinitely. In order to solve this we chose to estimate resting heart rate at 65 Beats per Minute (BPM). At each window without a heart rate, we would calculate the missing value using (4) if the previous heart rate was above the estimated mean and (5) if the previous heart rate was below the estimated mean.

$$HR_t = HR_{t-1} - \frac{HR_{t-1} - \mu_{HR}}{2} \tag{4}$$

$$HR_t = HR_{t-1} + \frac{\mu_{HR} - HR_{t-1}}{2} \tag{5}$$

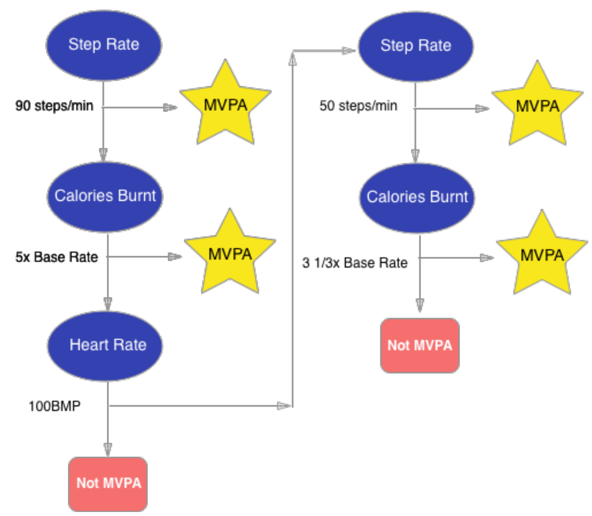

Fig. 6. MVPA Decision Tree Classifier

Once the missing values are filled in, we can actually begin classifying MVPA. We considered many different approaches for this classifier. Automated classifiers built into WEKA had poor recall, notably a high false negative rate. Logit and Probit regressions did not yield good results either; Betas were sporadic, did not make much sense, and had low t-scores. We found the best results with a hand built decision tree, shown in figure 6. We realized that step rate was our most informative feature with calories expended being our next most informative feature. Thus, we set thresholds on step rate and calories burnt and looked at both of those data points. If neither of these thresholds were met, we looked a heart rate. We found that heart rate was not good enough to classify MVPA alone. The readings given by the smartwatch sensors were not always accurate, and even if they were, heart rate alone is not a good determinant of MVPA. This is because somebody could have an elevated heart rate from excitement, stress, or even caffeine. That being said, we determined that heart rate was still a feature worth looking at. If neither step rate or calories burnt met the threshold, we checked heart rate. If avg, max, or min heart rate was above 100 BPM, we relaxed our constraints on Step Rate and Calories Expended. This allowed heart rate to affect the classification without giving it a final say.

We formed and evaluated this scheme using Google Fit data from members of the group. We each spent 1-2 hours conducting a fitness trial. These trials consisted of MVPA: running and walking briskly, and non MVPA: walking casually, sitting, standing, and performing household chores. We labeled the time windows that we performed each activity before. This classifier worked very well and was able to distinguish a brisk walk from a casual walk.

IV. METHODS

A. EXPERIMENTAL SETUP

As mentioned above, our experiment targets participants in the transition from the action stage of the trans-theoretical model to the maintenance stage. However, due to time limitations, our study was conducted over the period of two weeks. Since this time frame is inadequate for testing whether a participant will be able to maintain a fitness plan, we focused on testing participants in the action stage. We validated this using a pre-test survey to gauge a participant's level of commitment to attaining higher levels of physical activity. That being said, our application is designed with users in the maintenance stage in mind as well. GambleFit is based on mechanisms that support continued short-term benefits (Acticoins) and encourage habit-forming (by teaching users how to control fitness plans with attainable goals). Therefore, we believe it is applicable to both stages equally - although further studies will be needed to validate this claim.

All participants were required to have a basic familiarity with smartwatches and wearable activity trackers in order to minimize the learning curve for the experiment. This familiarity was also determined by our pre-test survey, which was necessary for our study because of our limited sample size and timeline to eliminate user error.

Because of the limited timescale and the preliminary stage of prototype, our study aims to mostly test the system and its ability to report accurately over a study period. Thus, our study consists of 4 participants randomly divided into two groups. One of these groups is the test group and the other is the control group. Both of these groups use the same system (described above) for obtaining fitness data in terms of minutes of MVPA per day. This is done in an attempt to eliminate errors in our minutes of MVPA calculation. The changes between each group appear in the interface presented to users. The test group receives the full version of the GambleFit application (see framework in section III) with daily adjustable goals through the gambling mechanism. The control group receives an interface that only reports a static MVPA figure and a daily minutes of MVPA goal. Goals are suggested when the app is started which are based on the published physical activity guidelines for Americans [11]. However, these recommended goals are not in any way adjusted to the specific user, because we want our users to learn how to set their own goals.

Once testers received their tracking devices, they were trained on how to use their assigned application and smartwatch. Figure 7 shows the two components of the system that were used by our testers. As described in the system section, a smartphone was used to collect data from Google Fit and the smartwatch. Testers were instructed to wear the smartwatch on one of their wrists and to keep the smartphone on their person, either inside a purse or inside a pant pocket. Testers were then given two weeks to track their activity and

Fig. 7. Sensor Locations

reach their goals. The test group was able to update their day-to-day goals and were evaluated against their initial goal and how accurately they set daily goals.

At the end of the study, smartwatches and smartphones were collected and the data was analyzed from the back-end servers. Testers were also given a post-test survey. This survey contained questions analyzing the self-reported effectiveness and usability of the application. We factor these measurements into the overall effectiveness of the GambleFit system (from both a behavioral and usability perspective).

B. MEASURMENTS AND EVALUATION

Our system is evaluated on two principles that represent usability of the application and the effectiveness of the application at encouraging behavior change. The first is the level of engagement we see with the application. The second is the level of physical activity we see in participants. The engagement level is important in analyzing whether the system is appealing to use and whether its usability factors into the efficacy of the system. It provides a way to understand the importance of usability in fitness tracking applications. The measurements and evaluation on physical fitness provide us with the effectiveness of the system in encouraging behavior change in general.

1) Engagement: To measure the level of engagement with the application we looked at two heuristics. First, through the use of application usage time, we actively measured how regularly the user interacted with our game. Secondly, we looked at the satisfaction users had with the game to gauge interest and usability levels that will sustain long-term use of the application. This was performed using a questionnaire based on the Mobile Application Rating Scale as well as the study on Mobile User Engagement by Kim et. al. [9]. We combine this with the post-test results to see what level of usage (or lack thereof) is attributable to the user interface from a human-computer interaction perspective.

2) Physical Activity: Since we targeted users in the action stage of transtheoretical model, we adhere to the guidelines set forth in the 2008 Physical Activity Guidelines for Americans. [11] In general, these guidelines suggest at least 150 minutes of moderate-intensity, 75 minutes of vigorous-intensity, or an equivalent combination of moderate to vigorous-intensity activity (MVPA). Thus, our system uses MVPA as the basic metric for tracking fitness level and activity.

As mentioned above, minutes of MVPA (mMVPA) can not be collected from any sensors. Instead we use the features described above (step count, calories expended, and heart rate) to obtain mMVPA based on a decision tree that analyzes 5 minute buckets of fitness activity. This decision tree acts as a basic classifier that converts the above features as either being part of MVPA or not at a 5 minute granularity.

Heart rate data is collected using a third party app and stored in the Google Fitness Store. Calories expended and step count data are obtained directly from Google Fit (available on Android smartwatches and smartwear).

For each participant, we used the mean and median of their daily difference between achieved MVPA and target MVPA. Positive values mean they beat their goal and negative values means they did not obtain their goal. We then found the mean and median of each of these values for each participant. Then with we did an unpaired T-Test for the mean difference between each group.

We also investigated daily MVPA achievement. For each user, we caluclated their average daily MVPA amount. Then, we calculated the average MVPA achieved for the group, using each individual's average. We did this using both means and medians to account for outliers. We conducted unpaired T-Tests for the mean differences between each group.

3) Evaluation: The efficacy of the mMVPA classifier is be assessed using a standard confusion matrix. All data is analyzed using standard statistical analysis. These statistics will be used to ultimately test whether our hypothesis holds true.

V. RESULTS

1) Engagement: We aggregated the data of each group and conducted an unpaired T-test on the means of self-reported time spent on the application. The mean engagement metric for the control group was 5.23s with a standard deviation of 0.44s. The mean engagement metric for the test group was 16.33s with a standard deviation of 7.93s. Thus the difference in means is 11.10s. Our pooled standard deviation is 5.03s, giving us a standard error of 1.07s. Finally this gives us a T-Statistic of 10.37. This shows that engagement time improved using GambleFit with a confidence of 99% and p-value $<.00001$.

We also analyzed the post-test questionnaire from the users who had GambleFit to understand the usability of the full application, which would drive long-term engagement.On a scale of 1 - 10, GambleFit users reported an average score of 5.5 on whether they would choose our application over a regular Google Fit application. They reported average scores below 5 on the same scale for questions on how interesting or fun the application was, although they also reported an average score of 6 for a question on whether the application ultimately helped them improve their activity levels. In the qualitative feedback, all users mentioned the lack of detail in activity displayed as well as inability to access social functions as the main disadvantages of the application. No users were able to recognize the game's purpose in making them more cognizant of their daily activities.

2) Physical Activity: Throughout the study, we recorded the fitness activity of each user in the Google Fitness Store. After daily analysis during check-in time, we recorded desired features and calculated MVPA in a separate database for analysis. After our short, preliminary study, we were able to aggregate this data and report on total daily steps, calories expended, and MVPA over the course of the study for each of our users. Figure 8 shows user step counts over the course of the study. Note that User 1 and User 4 were part of the test group given the GambleFit app, and User 2 and User 3 were given the control app. Also note that there are gaps in reported data largely due to users missing check-ins. Note that User 4 has no data past March 2 due to compatibility issues with their phone and our application. We believe such issues arose from fragmentation within the Android ecosystem, and our application has been adjusted to better accommodate a variety of devices. Figure 9 shows a similar graph for daily calorie consumption. Finally, figure 10 shows the overall trend of minutes of MVPA per user over the course of our study.

For each participant we took the mean of their daily difference between achieved MVPA and target MPVA. Positive values indicate that the user beat their goal and negative values indicate that user did not obtain their goal. We then found the mean of each of these values for each participant. A summary of our findings broken down by user (User 4 is omitted due to incomplete data) is given in table I. This is consolidated using an unpaired t-test in table II. The t value given was 2.8 with 20 degrees of freedom. This equates to a

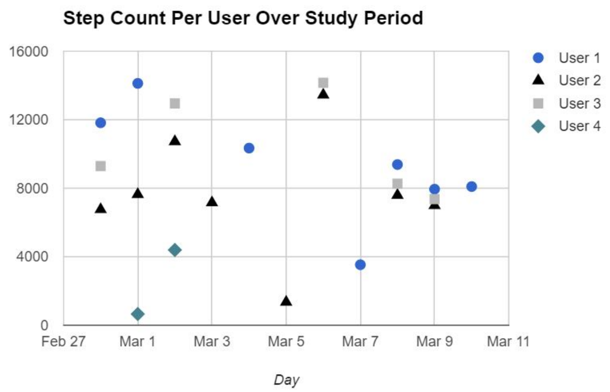

Fig. 8. Test Group: User 1, User 4 - Control Group: User 2, User 3

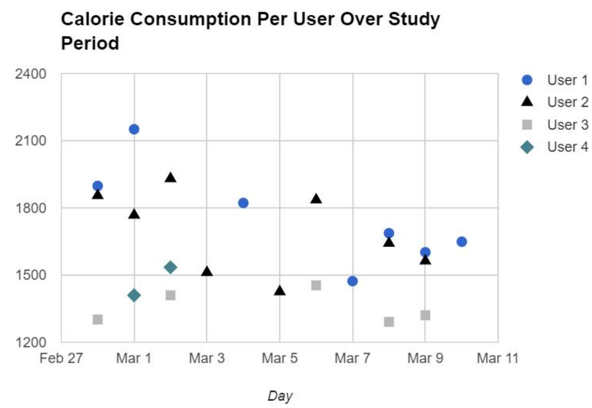

Fig. 9. Test Group: User 1, User 4 - Control Group: User 2, User 3

P value of .0101. The standard error of difference is 14.317.

	User 1	User 2	User 3
Mean Difference	-6.8	64.3	3
Difference Std Dev.	17.8	26.3	20.79

TABLE I
COMPARISON OF DAILY GOAL TO ACTUAL MINUTES OF MVPA

	Control Group	Test Group
Mean of Difference	6.8	-33.7
SEM	7.08	10.80
SD of Difference	17.8	23.6
N	9	13

TABLE II
COMPARISON OF DAILY GOAL TO ACTUAL MINUTES OF MVPA BY GROUP

Finally, for each group we found the average daily amount of MVPA for each user. Then we found the average of averages for each group, giving us a group average. We did this with both means and medians to see if there is an effect of outliers. This data is presented in table III. Then we conducted an unpaired T-Test for the mean difference between each group as well as a paired T-Test for the median

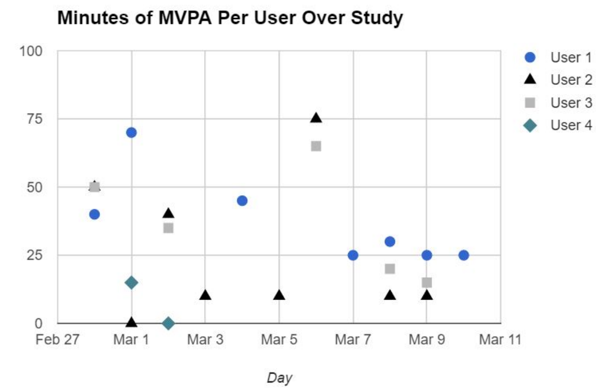

Fig. 10. Test Group: User 1, User 4 - Control Group: User 2, User 3

difference between each group.

	Control	Test
Mean/Participant	31.31	14.82
Median/Participant	22.50	18.75
σ_{mean}	8.04	20.96
σ_{median}	17.68	15.91

TABLE III
DAILY MVPA FOR EACH GROUP

We use these daily MVPA values to conduct our unpaired T-Test to see if the application had a statistically significant impact on each users physical activity. Using the medians we found a T-Score of 0.223. Using the means we found a T-Score of 1.039.

For each participant we calculated their mean and median difference from their goal each day.

VI. DISCUSSION

When it comes to engagement, the results show that users of GambleFit spend significantly more time on the application than the control users. This is to be expected, since GambleFit players have more tasks to complete, such as guessing and wagering, than control users, who only have to view their activity levels. Users, however, could not recognize the application's purpose in making them more cognizant of the daily activities. Despite this, they saw the application improving their daily activities, only because it encouraged them to be more accountable to their goals. This suggests that the guessing component of the game is insufficient in encouraging GambleFit users to consider the activities they had done that day and how they would equate to MVPA levels.

Furthermore, with the low rating of the game in the fields of fun and interest, it is improbable that users will engage in the game at a sustainable level, which was one of our goals. This suggests that merely gaining or losing Acticoins is not a high stake activity, and the inclusion of features such as social comparisons with peers or redemption opportunities at local restaurants may make the game more fun and interesting. Overall, GambleFit's engagement is not as high as we wanted for it to be a successful application, and requires modifications as noted.

As mentioned above, the study was limited in both time and participants, and thus serves mostly as a preliminary test of our system in a study environment. However, even with our limitations, we were still able to see some important trends in our data. The most interesting thing to note is the general drop off in activity level from the beginning of our study to the end. The mean minutes of MVPA for all study participants during the first two days of the study was 32.5 minutes. The mean minutes of MVPA for all study participants during the last two days of the study was 17.5 minutes. Even in our short study period we see all users dropping off in their commitment to physical activity. Naturally, this trend is mirrored in the features we use to determine MVPA; both step count and calorie consumption also follow a general downward trend.

The goal of our application is to mitigate the complete drop off activity levels in the long term. Although this is harder to see in our short study, we can guess whether or not our application help will mitigate a full drop off in fitness levels by looking at the goal setting habits of our users. Based on our research into theory, we believe that our application is more likely to keep users on track to remain physically active because they are learning how to set attainable fitness goals. As seen in table I and found in the t-test referenced by table II, the test group is able to more closely achieve their daily MVPA goals without under or over estimating to the same degree as the control group in a statistically significant manner (The P value is .0101).

The T-Tests are clearly not significant for MVPA for each group. Since we only had 2 particpants in each group and 1 degree of freedom for our 2-tailed t-test, the required T-score for significance is 12.71. As these are very far from this range, we cannot draw any conclusions here. A trial with more users would be very important in futher investigating the effects of this application. With a small sample size like this, it is very difficult to detect a marginal increas in activity.

	Predicted	
	Positive	Negative
Positive	20	2
Negative	4	586

TABLE IV

Confusion Matrix for MVPA Classifier

Precision	.833
Recall	.909
Accuracy	.990

TABLE V

Precision, recall, and accuracy of the confusion matrix in Table IV

Table IV shows the confusion matrix for our MVPA Classifier. Based on the precision, recall, and accuracy calculations shown in Table V, we believe that the classifier performed well. The results given by our classifier were much better than anything we could obtain using Weka or logit/probit regressions. Step count was the most informative feature, closely followed by calorie count. Heart rate was not incredibly informative, but was still important to consider in the classification.

VII. Conclusion

In this paper, we present GambleFit, a mobile application designed to help people become more aware of their physical activity levels. Every night, GambleFit requires players to check-in with the application and guess how many minutes of MVPA they engaged in that day. After guessing, the application shows the user how many minutes of MVPA they actually engaged in. GambleFit rewards players with Acticoins based on how well they estimate their activity and it encourages them to set activity goals for themselves each day.

To measure user activity levels, GambleFit obtains fitness data from Google Fit. This data, which includes step count, calories expended, and heart rate, is then processed using a decision tree to determine if the user was actually engaged in MVPA for a given time interval. These calculations are performed in both GambleFit and the control application, and they are used to determine the user's minutes of MVPA for the day.

One of the limitations of this project was the short time frame. Having more time to spend testing the application and collecting data from the users would have helped us to conduct a more rigorous evaluation of GambleFit. Additionally, our sample size was limited to four people, which resulted in our test group and control group having only two members each. With additional time and devices, we may have been able to determine the efficacy of our approach more accurately.

In the future, there are several changes that can be made to GambleFit to improve its performance. First, it would be useful to have the application send push notifications at certain points in the day. For example, it could remind the user every morning what their goal is, and it could remind the user to check-in with the application at nighttime. Additionally, finding a way to attach an actual value to the Acticoins would potentially make the application much more successful. One possible way would be to partner with socially-minded businesses, like Starbucks or Whole Foods, and have them offer material rewards for a certain number of Acticoins.

Acknowledgments

The GambleFit team would like to thank the following individuals and organizations for making this project possible:

- Northwestern EECS, McCormick Engineering
- HABits Lab, Prof. Nabil Alshurafa
- Shibo Zhang
- Dr. Amanda Paluch
- Dr. Bonnie Spring

References

[1] Alshurafa, Nabil, et al. ”Battery optimization in smartphones for remote health monitoring systems to enhance user adherence.” Proceedings of the 7th international conference on Pervasive Technologies Related to Assistive Environments. ACM, 2014.

[2] Alshurafa, Nabil, et al. ”Designing a robust activity recognition framework for health and exergaming using wearable sensors.” IEEE Journal of Biomedical and Health Informatics 18.5 (2014): 1636-1646.

[3] ”A SMART Guide to Goal Setting.” ACE Fit. American Council on Exercise, n.d. Web. 16 Feb. 2017.

[4] Bandura, Albert. ”Social Cognitive Theory of Self-regulation.” Organizational Behavior and Human Decision Processes 50.2 (1991): 248-87. Web.

[5] Direito, Artur, Yannan Jiang, Robyn Whittaker, and Ralph Maddison. ”Apps for IMproving FITness and Increasing Physical Activity Among Young People: The AIMFIT Pragmatic Randomized Controlled Trial.” Journal of Medical Internet Research 17.8 (2015): n. pag. Web.

[6] Douris, Peter C., et al. ”Comparison between Nintendo Wii Fit aerobics and traditional aerobic exercise in sedentary young adults.” The Journal of Strength & Conditioning Research 26.4 (2012): 1052-1057.

[7] ”Fitbit.” Fitbit Official Site for Activity Trackers & More. N.p., n.d. Web. 16 Feb. 2017.

[8] Jakicic, John M., et al. ”Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial.” Jama 316.11 (2016): 1161-1171.

[9] Kim, Young Hoon, Dan J. Kim, and Kathy Wachter. ”A Study of Mobile User Engagement (MoEN): Engagement Motivations, Perceived Value, Satisfaction, and Continued Engagement Intention.” Decision Support Systems 56 (2013): 361-70. Web.

[10] Nutbeam, Don, Elizabeth Harris, and Marilyn Wise. Theory in a Nutshell: A Practical Guide to Health Promotion Theories. Sydney: McGraw-Hill, 2014. Print.

[11] ”Physical Activity Guidelines for Americans Summary.” Physical Activity Guidelines for Americans. Health.gov, 2008. Web. 16 Feb. 2017.

[12] Tabak, Monique, Marit Dekker-Van Weering, Hylke Van Dijk, and Miriam Vollenbroek-Hutten. ”Promoting Daily Physical Activity by Means of Mobile Gaming: A Review of the State of the Art.” Games for Health Journal 4.6 (2015): 460-69. Web.

[13] Yamada, Yosuke, et al. ”Light-intensity activities are important for estimating physical activity energy expenditure using uniaxial and triaxial accelerometers.” European journal of applied physiology 105.1 (2009): 141-152.