The Social fMRI: Measuring, Understanding, and Designing Social Mechanisms in the Real World

Nadav Aharony[1], **Wei Pan**[1], **Cory Ip**[1], **Inas Khayal**[1,2], **Alex Pentland**[1]
{nadav, panwei, coryip, ikhayal, pentland}@media.mit.edu
[1]The Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Masdar Institute of Science and Technology, Masdar City, Abu Dhabi.

ABSTRACT
A key challenge of data-driven social science is the gathering of high quality multi-dimensional datasets. A second challenge relates to design and execution of structured experimental interventions in-situ, in a way comparable to the reliability and intentionality of ex-situ laboratory experiments. In this paper we introduce the *Friends and Family* study, in which a young-family residential community is transformed into a living laboratory. We employ a ubiquitous computing approach that combines extremely rich data collection in terms of signals, dimensionality, and throughput, together with the ability to conduct targeted experimental interventions with study populations. We present our mobile-phone-based social and behavioral sensing system, which has been deployed for over a year now. Finally, we describe a novel tailored intervention aimed at increasing physical activity in the subject population. Results demonstrate the value of social factors for motivation and adherence, and allow us to quantify the contribution of different incentive mechanisms.

Author Keywords
Social Computing, Mobile Sensing, Social Health

ACM Classification Keywords
H.4.m Information Systems: Miscellaneous; J.9.d Mobile Applications: Pervasive computing

General Terms
Algorithms, Experimentation, Measurement

INTRODUCTION
Imagine the ability to place an imaging chamber around an entire community. Imagine the ability to record and display nearly every facet and dimension of behavior, communication, and social interaction among the members of said community. Moreover, envision being able to conduct interventions in the community, while measuring their effect - by both automatic sensor tools as well as qualitative assessment of the individual subjects. Now, think about doing this for an entire year, while the members of the community go about their everyday lives.

Utilizing ubiquitous computing devices and methodologies, we developed a mobile-phone-centric social and behavioral sensing system that we have deployed with 130 adult members of a young-family living community for over a year now. During this year we have amassed what is, to the best of our knowledge, an unprecedented longitudinal dataset, which we dub the *Friends and Family* dataset. The dataset includes continuous collection of over 25 phone-based signals - including location, accelerometery, Bluetooth-based device proximity, communication activities, installed applications, currently running applications, multimedia and file system information, and additional data generated by our experimental applications. In addition, we collect financial information through receipts and credit card statements, logging of Facebook socialization activities, daily polling of mood, stress, sleep, productivity, and socialization, as well as other health and wellness related information, standard psychological scales like personality tests, and many other types of manually entered data by the participants.

The data enable us to construct multiple network modalities of the community - such as the phone communication network, physical face-to-face encounters network, online social network, self-reported network, and more. We use these networks to investigate questions like how things spread in the community, or how communication activities predict the spread of mobile applications. Another example using both the individuals financial status and social behaviors is understanding the causality question raised by Eagle et al. [15], where we discovered that the causality may go along the opposite direction. However, in the current paper we direct the discussion to an additional aspect of the study, which is the design and execution of experimental interventions while measuring their effect on individual and group behavior.

Out of several interventions conducted over the past year and planned for the upcoming months, in this report we focus on a fitness and physical activity intervention conducted between October to December of 2010. Using an experimental intervention within the Friends and Family study population, we test social mechanism-design principles. In particular, we propose a novel social mechanism in which subjects are rewarded based on their peers' performance and not their

UbiComp '11, Sep 17-Sep 21, 2011, Beijing, China.

own. Results suggest that: (1) Social factors have an effect on the physical activity behavior, motivation, and adherence over time. (2) Social incentives, and particularly our novel Peer-Reward mechanism encouraging social influence among participants, support higher activity returns per dollar invested in the system. (3) Finally, results support the notion of a complex contagion [10] like effect related to pre-existing social ties between participants.

The contributions of the work described in this paper are threefold: We present the Friends and Family study and very rich dataset; We describe the field-proven system that has been deployed and used for over a year, which we intend to release as an open source platform for social and behavioral data collection and feedback; We conducted the fitness intervention and find results that contribute to our understanding of social incentives and motivation in real-world contexts.

In the remainder of this paper, we first review related literature and the contextual underpinning of our proposed vision. We then introduce our approach. In the next section we go into our methodology - the Friends and Family living laboratory and its characteristics. Next we describe our system architecture, and then review the experimental design of our social fitness intervention. We emphasize how our gained familiarity with the community comes into play in the study design. Finally we analyze key results of the intervention.

RELATED WORK AND CONTEXT

Ubiquitous Social Observatories

In recent years the social sciences have been undergoing a digital revolution, heralded by the emerging field of "computational social science". Lazer, Pentland et al. [24]. describe the potential of computational social science to increase our knowledge of individuals, groups, and societies, with an unprecedented breadth, depth, and scale. [24] highlights challenges in terms of scientific approach for observation and intervention when dealing with real people in their living environments, including issues of subject privacy, monitoring, and altering of environments during the discovery process.

Figure 1 gives a high-level qualitative overview of social observatories and datasets, comparing them along axes of sample size, duration, and a very rough notion of "throughput" or the amount of information in the datasets. The idea is that dataset throughput is a function of the data dimensionality (number of different signals collected), its resolution (e.g. raw or aggregate), sampling rate (how often data is collected), and unique information in it (an accelerometer sensor lying on a desk for a week does not collect a lot of information). This diagram illustrates the potential of ubiquitous technologies for the design of social observatories and the collection of very rich datasets. At the bottom of the diagram are traditional sociology studies as well as many of the corporate "donated" datasets. Leading traditional dataset include, for example, the Framingham Heart Study [13], which stands out for its duration and a subject pool of several thousands, however its "throughput" is low as subjects were sampled approximately once in three years.

The pervasiveness of mobile phones has made them ubiquitous social sensors of location, proximity and communications. Because of this, mobile phone records from telecom companies have proven to be quite valuable in particular. Gonzales et al. show that cell-tower location information can be used to characterize human mobility and that humans follow simple reproducible mobility patterns [21]. Eagle et al. find that the diversity of individuals relationships is strongly correlated with the economic development of communities [15]. These and other corporate "donated" datasets are usually characterized by having, on one hand, information on very large numbers of subjects, but on the other hand, this information is constrained to a specific domain (email messages, financial transactions, etc.), and there is very little if any contextual information on the subjects themselves. This is why, although their sample size may be in the millions, they are relatively low on the throughput axis. As example, [22], each sampling point was an aggregated 15-day call summary of anonymous phone users. In addition, domain-limited results might be harder to generalize for the physical world, as discussed by Onnela et al. in context of Facebook data[32]. Finally, there is the offline nature of most existing datasets which are based on previously collected data, making it harder to test cause and effect using these datasets. Yet these datasets are yielding significant contributions to our understanding of society, one might imagine that by increasing the dimensionality and throughput, such datasets could lead to even further advancement.

An alternative approach is a bottom-up one, collecting data at the level of the individual. Eagle and Pentland [16] defined the term"Reality Mining" to describe collection of sensor data pertaining to human social behavior. They show that using call records, cellular-tower IDs, and Bluetooth proximity logs, collected via mobile phones , the subjects' social network can be accurately detected, as well as regular patterns in daily activity [16, 17]. This initial study was then expanded in Madan et al.[27], who conducted a similar experiment and show that mobile social sensing can be used for measuring and predicting the health status of individuals based on mobility and communication patterns. They also investigate the spread of political opinion within the community [28]. Other examples for using mobile phones for social sensing were done by Montoliu et al. [30] and Lu et al.[26]. Most of these were of an observational nature, and have not performed controlled experimental interventions for exploring social mechanism. Other types of sensor-based "social observatories" are the Sociometric Badges by Olguin et al. that capture human activity and socialization patterns via a wearable sensor badge [31]. A key aspect of the Sociable Badges is that they have been deployed in studies where sensor feedback was given to the corporate participants [31].

Physical Activity Sensing and Feedback

In this paper we focus on a specific problem from the domain of health and wellness: Studies have shown a great increase in obesity and related chronic medical conditions over the last several decades. Physical activity has been shown to help alleviate the burden of obesity and other health conditions [6, 8]. Over the past two decades, the accelerometer

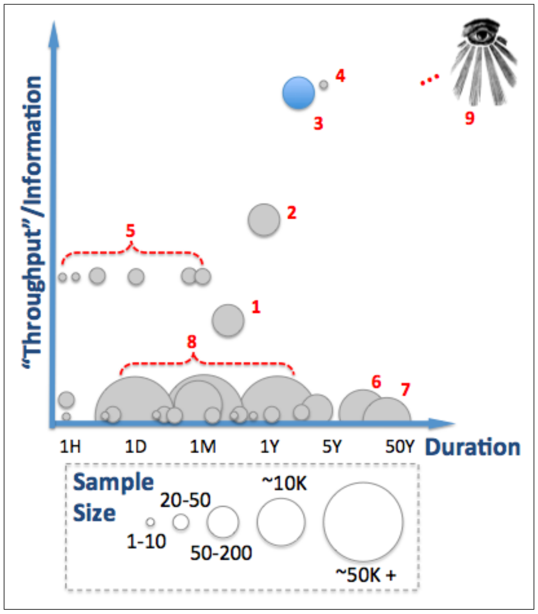

Figure 1. Qualitative Overview of social science "observatories" and datasets, along axes of data collection duration, qualitative"throughput" , and the size of the subject sample.(1)Reality Mining[16], (2) Social Evolution[28], (3) Friends and Family Dataset, (4) rich-data pioneers [3, 20] (5) Sociometric Badge studies [31], (6) Midwest Field Station [5], (7) Framingham Heart Study[13], (8) Large Call Record Datasets [21, 15, 22] , (9) "Omniscient"/All-Seeing View

has been established and refined as a tool for tracking physical activity [9, 18, 35]. Accelerometry-based sensors have been found to provide more accurate estimates than other widely-used proxies for energy expenditure [18]. Although there is some error associated with using accelerometers to track energy expenditure in free-living situations, a significant relationship between accelerometer output and energy expenditure has nevertheless been established [9]. Several studies in the ubiquitous computing literature have targeted this important problem domain. Ubifit [12] is one of the most extensive works investigating ways to encourage physical activity. Other projects include Fish'n'Steps [25] and Houston [11], among others. These works investigate diverse aspects of the problem, such as user interface, goal setting, or techniques for using the accelerometer for accurate expenditure measurement and activity detection.

Of particular relevance are those studies that involve social components [25, 11, 4, 34, 19]. It has long been established that social support is a resource for behavioral change an and indicator for health [7], however here is still much to be learned about the fine-grained *social mechanisms* related to physical activity behavior, as well as how to leverage such insights in *designing better socially-aware interventions and mechanisms* for encouraging healthy behavior change.

For activity measurement, relevant works are and those using unaugmented phone-based activity detection [4, 33], whereas the majority of studies to date used additional measurement devices that need to be carried by subjects. In the consumer world, a growing number of activity measuring mobile applications such as CardioTrainer [1] use the phone's acclerometer, combined with visualization and other feedback to help users increase their physical activity levels. Most applications aim to provide a step count measurement, and ask the users to hold the phone in a certain orientation while exercising in order to deliver accurate measurements.

THE SOCIAL fMRI

In the medical realm, Magnetic Resonance Imaging, MRI, is considered one of the most comprehensive diagnostic techniques available, and functional MRI, fMRI, is one of the leading tools used for studying the brain through response to carefully designed stimuli. Analogously, we define Social Mechanism-design and Relationship Imaging, or Social MRI, which allows detailed sensing and imaging of social systems through the use of mobile phones, credit cards, social media, and telecommunications for social and behavioral sensing platform. *Social fMRI* takes it a step further - allowing for specifically designed stimuli and interventions to the social system, while measuring the individual and collective response. Just as fMRI helps us understand the inner workings of the brain, we hope that the Social fMRI approach could help us understand the inner workings of social systems and the way humans interact and react to one another. More than just an aspiration, in this paper we show a proof of concept as to how this could be done.

The general framework of the Social fMRI idea is a combination of a longitudinal living-laboratory/social-observatory type of study, coupled with a supporting system infrastructure that enables the sensing and data collection, data processing, and also a set of tools for feedback and communication with the subject population. The Social fMRI implements and extends the ideas of the Reality Mining approach [16], by (1) adding much greater data richness and dimensionality, combined with (2) a strong element of active interaction and carefully designed experimental stimulation of the study population.

METHODOLOGY

Living Laboratory: The "Friends and Family" Community

Starting March 2010, we initiated a living laboratory study conducted with members of a young-family residential living community adjacent to a major research university in North America. All members of the community are couples, and at least one of the members is affiliated with the university. The community is composed of over 400 residents, approximately half of which have children. The residence has a vibrant community life and many ties of friendship between its members. We shall refer to this residence as the "*Friends and Family*" community.

This study involves a relatively different subject population when compared to previous ubiquitous computing observatory studies. For example, colleagues and co-workers in Reality Mining [16], and undergraduates in [27]. The Friends and Family community includes a much more heterogeneous subject pool, and provides a unique perspective into a phase in life that has not been traditionally studied in the field of ubiquitous computing - married couples and young families.

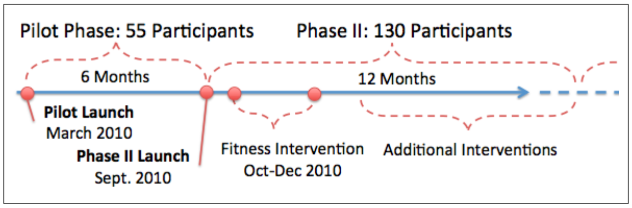

Figure 2. High level timeline for the Friends and Family study.

As depicted in Figure 2, a pilot phase of 55 participants launched in March 2010. In September 2010 phase two of the study included 130 participants, approximately 64 families. Participants were selected out of approximately 200 applicants, in a way that would achieve a representative sample of the community and sub-communities. One of the reasons for keeping the number below 150 is that these numbers fit well with Dunbar's social evolutionary theory regarding the number of people humans are able to maintain a relationship with [14]. Throughout the study we ask about social closeness between all participants in the study, and numbers larger than Dunbar's number could become quite tedious. We refer to experiments in our scale as "Dunbar scale" experiments. The research goals of the longitudinal study touch on many aspects of life, from better understanding of social dynamics to health to purchasing behavior to community organization. The two high-level themes that unify these varied aspects are: (a) how people make decisions, with emphasis on the social aspects involved, and (b) how we can empower people to make better decisions using personal and social tools.

Study Data Collection

One of the key goals of the Social fMRI idea is the collection of multi-modal and highly diverse range of signals from the subject population. We wanted to gather data on numerous network modalities, so that their properties and interrelation could be better understood. We applied a user centric, bottom up approach utilizing the following components:

Mobile Phone Sensing Platform

This is the core of the studys data collection. Android OS based mobile phones are used as in-situ social sensors to map users activity features, proximity networks, media consumption, and behavior diffusion patterns. The mobile phone platform is described in more detail in the next section. We did not sponsor phone plans or data plans - users received a mobile phone that fit their desired provider, and they were responsible to port their existing account to it or open a new account. The condition was that the study phone be their primary phone for the duration of the study.

Surveys

Subjects complete surveys at regular intervals, combining web-based and on-phone surveys. Monthly surveys include questions about self perception of relationships, group affiliation, and interactions, and also standard scales like the Big-Five personality test [23]. Daily surveys include questions like mood, sleep, and other activity logging.

Purchasing Behavior

Information on purchases is collected through receipts and credit card statements submitted at the participants discretion. This component targets categories that might be influenced by peers, like entertainment and dining choices.

Facebook Data Collection Application

Participants could opt to install a Facebook application that logs information on their online-social network and communication activities. About 70% of subjects opted to install.

Subject Protection and Privacy Considerations

The study was approved by the Institutional Review Board (IRB) and conducted under strict protocol guidelines. One of the key concerns in the design of the study was the protection of participant privacy and sensitive information. For example, data is linked to coded identifiers for participants and not their real world personal identifiers. All human-readable text, like phone numbers and text messages are captured as hashed identifiers, and never saved in clear text. Collected data is physically secured and de-identified before being used for aggregate analysis. An second important consideration was for being as unobtrusive as possible to the subject's life routines. Participants are able to keep the phone at the end of the study. For mandatory out-of-routine behavior that asked of participants, like filling out surveys, subjects are compensated (e.g. $10 for completing the monthly survey). Participation in intervention or sub-experiment on top of the main study components is completely optional, and interventions are carefully designed with the interests of the participants in mind.

SYSTEM ARCHITECTURE

Mobile Phone Platform

The phones run our software platform, which periodically senses and records information such as cell tower ID, wireless LAN IDs; proximity to nearby phones and other Bluetooth devices; accelerometer and compass data; call and SMS logs; statistics on installed phone applications, running applications, media files, general phone usage; and other accessible information. Over 25 different types of data signals are currently collected. The system also supports integration of user-level apps, like an alarm clock app we developed, for additional data collection and interventions. The phone system also has a survey application. Sample screenshots can be seen in Figure 3. The configuration is set so that battery-intensive actions (e.g. GPS scans) are performed in intervals allowing usefulness while minimizing battery drain. A remote configuration capability allows for fine-tuning the system, with a goal of enabling a minimum of 16 hours between charges. We are working towards releasing the software, named "Funf", as an open source framework [2].

Data Formats and Server Communications

Phone data is is saved in SQLite file format, with files rotated every three hours to reduce data loss due to file corruption and to allow periodic uploading to the back-end. Since many participants do not have a mobile data service plan, the system was designed in a "delay tolerant" way: In the absence of network access, the phone accumulates the collected database files locally. Once server connection is made

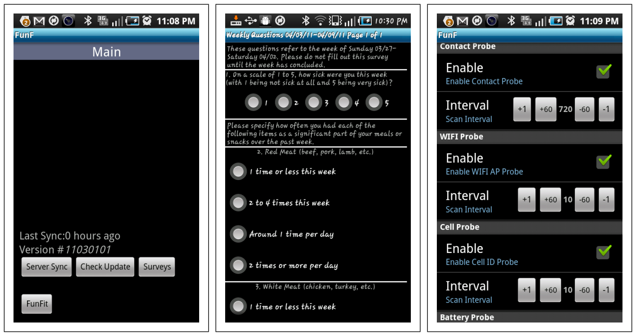

Figure 3. Sample screenshots: Sync-state and version display (left), survey (center), and probe preferences debug screen

(for example a participant connects to WiFi to browse the web), the system attempts to upload files. Once uploaded, files are also encrypted and saved in a backup directory on the phone, until data integrity is confirmed. In addition, the system downloads from the server parameter configurations files that define which data signals to collect, how often, and additional configuration parameters. It also checks for new versions of itself and notifies the user to update when new versions are available. The system downloads any new surveys that should be presented to the user.

Back-end

The server-side back-end processes all incoming SQLite files and insets them into a central MySQL database. It sends email reports to investigators about status of phones and alerts of any issues. Additional services provide data for interventions and personal data visualization for participants.

FITNESS-CENTERED SOCIAL INTERVENTION DESIGN

Between October-December 2010, an active intervention was carried out with the Friends and Family study pool. Its main goal was to explore the question of understanding social influence and motivation in the context of health and wellness activities. The intervention was presented to participants as a wellness game to help them increase their daily activity levels. 108 out of 123 active subjects at the time elected to participate. Subjects were divided into three experimental conditions: Control, Peer-View, and Peer-Reward (described below), allowing us to isolate different incentive mechanisms related to monetary reward, the value of social information, and social pressure/influence. Following an initial period where baseline activity levels were collected, all intervention subjects were given feedback on their performance in the form of a monetary reward, R, which was calculated as a function of their activity. Reward of up to $5 was allocated every three days. Participants were presented current, past, and total reward accumulated, as shown in Figure 4, and could access their reward page via browser or the phone. Each group received a variant adapted to its condition. The game was not designed as a competition, and every subject had the potential to earn the maximal reward.

Experimental Conditions

Control Condition

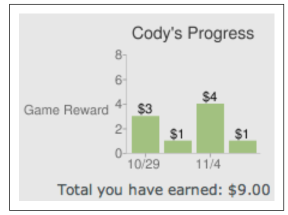

Figure 4. Reward display for participants in the Control condition

All conditions have a baseline of self-monitoring. In the control condition, subjects saw only their own progress as visualized in Figure 4. Also, reward given to the control subjects depended only on their own activity.

Experimental Condition 1: "Peer-View"

In the first experimental condition, "Peer-View", subjects were shown their own progress and the progress of two "Buddies" also in the same experimental group. In turn, the subject's progress was visible to two other peers in the same experimental group. Each subject's reward still depended on his own activity.

Experimental Condition 2: "Peer-Reward"

We propose a novel condition aimed at generating increased incentives for social influence and possibly the leveraging of social capital. In this "Peer-Reward" mechanism, subjects were shown their own progress as well as that of two Buddies, but this time subjects' rewards depended solely on the performance of their Buddies. At the same time, their own performance reward was split between two other peers, to which the current user was a Buddy. If subject A had Buddies B and C, the maximal reward A could receive for this period is still $5 per three-day period: $2.5 from B and $2.5 from C. The Peer-Reward feedback page displays how much reward they received from their Buddies, as well how much reward they are earning for the people they are Buddy to.

Accelerometer-based Activity Measurement

Our investigation of social mechanisms does not require accurate activity classification and step measurement. We decided to implement a less accurate but more robust algorithm for estimating activity levels, which allowed for increased flexibility in the way participants could carry the phone, and reduced the obtrusiveness of the study. Accelerometer scans were sampled in a duty cycle of 15 seconds every 2 minutes. During the 15 seconds, raw 3-axis accelerometer measurements are sampled at 5Hz rate and combined to compute the vector magnitude for each sample. The variance of the the magnitude in each one-second block is then computed [18]. The score was calculated by giving one point for every second, thresholded to three states 1)"still" 2)"moderate activity" 3)"high activity". However, in this paper we are interested in general change in activity levels and therefore combine the two active levels. Participants were not constrained in the way they should carry the phone. [33] found this did not interfere with activity measurement and classification, and our tests suggest this as well. Participants were told that the more they carry the phone on their person, the more of their activity would be accounted for their game score.

Game Reward Calculation

Game reward was calculated every three days, using a reference window of the seven days preceding the current 3-day bin. Average and variance for daily activity count are calculated for the reference window, as well the daily average for the current 3-day bin. Reward depended solely on an individual's performance, and was mapped to be between $0.50-$5, in $0.50 increments between one standard deviation above and below the reference average value. Values above or below the bounds received max or min value, respectively. To avoid discouragement due to zero reward, we did not give less than 50 cents per reward period.

Discussion of Experiment Design Considerations

One of the great advantages of the Social fMRI and other ubiquitous living-laboratory approaches is the ability to conduct interventions and structured experiments with the study population, as they live their everyday life. In contrast to most fitness-related studies who recruit participants specifically for the fitness study and many times pick participants who actively want to increase their physical activity [25, 11, 12], we faced similar challenges to those discussed in [29] for working with general populations in the wild. The sub-experiment had to be tailored to the nature of the subjects and the community, and be unobtrusive and attractive enough that the study population would want to opt-in.

We had to consider a range of attitudes towards physical activity. The intervention was thus designed as an non-competitive game, where each person is judged based on their own performance and performance change. A previously non-active participant could gain the same reward as a highly active one, while the highly active person would need to work harder. We also had to assume subjects might talk to each other and share information about the game. This is one of the reasons we made sure every participant would have potential to earn the same reward amount. Additional practical considerations included the fact that not everyone had data-plans, and data upload could be delayed. Since we needed it for the reward calculation, we added feedback to users about their data upload state, and also designed the accelerometer and reward three-day bins in a way that would allow for backlogged data to reach the server in time.

By creating a network structure rather than closed team structure for the social interventions (A receives reward for B and C's performance, while D and E receive reward for A's), we are able to disambiguate and focus on the diadic and asymmetric relationship of the person doing the activity vs. the person receiving the reward, motivated to convince the first.

PREPARATORY ANALYSIS

Self Reported Closeness

For the social conditions allocation, we wanted to leverage our knowledge of the subjects' network. We decided to use the network of self-perceived closeness since this network is explicit in participant's minds (as opposed to the Bluetooth collocation network, for example), and this was desirable for the experimental conditions. Each participant rated every other participant on a scale of 0-7,from 0(not familiar) to 7(very close). Basic analysis for the intervention participants network shows that it is a fully connected graph except one user. On average, each participant knows 14 other participants. Each participant has, on average, 7 explicit friendship ties (closeness > 2) in the study pool. The mean distance between any two participant is 1.9.

Experimental Condition Allocation

Based on this and marriage ties information, we designed an allocation algorithm to pair each participant in Peer-See and Peer-Reward with two buddies within their group. We wanted to ensure that at least some participants are paired with existing friends, while keeping the desired network topology and avoiding reciprocal pairings. Due to the sparsity of the friendship network, our division to disjoint experimental groups, and our enforced constraints, we formulated an integer programming optimization problem that attempts to prioritize closer friends as buddies with the following constraints: First, each participant should have exactly two Buddies. Second, participants cannot be their own Buddy. The third constrain prohibits two participants from being buddies of each other (reciprocity). Finally we prevented participants from having their spouses as buddies. This decision eliminates the unique and complicated effects resulting from marriage ties, and ensures that our fitness peer monitoring topology is purely constructed of regular friendship ties. The integer programming problem cannot be solved directly, and we applied an iterative approach: In each iteration, we randomized initial values and used the branch-and-bound algorithm to search for reasonable results, and we select the best solution among all iterations to match individual with their Buddies for both social condition groups in our experiment.

POST-INTERVENTION ANALYSIS

Subject Pool

Eleven subjects were removed from the study pool over the course of the intervention (due to prolonged technical issues that prevented reliable activity tracking, long durations of out of town travel, or dropping out of the longitudinal study entirely). Their data has been removed from the analysis, except for cases of analyzing peer effects for their Buddies. For details on the final number of subjects in each study condition, see Table 1.

Condition	Initial	Dropped	Total
Control	18	2	16
Peer-See	45	5	40
Peer-Reward	45	4	41

Table 1. Number of subjects in each condition

Intervention periods for analysis

For analysis of changes in activity level through the intervention, we divided the intervention into three periods: the baseline period before the beginning of the intervention was officially announced , the first 19 days of the intervention , and the second 20 days of the intervention . The periods are summarized in Table 2. For this analysis, the days after the intervention begins are broken up into two periods, and we focus on the latter one to account for any novelty effects and allow us to take a first look at the persistence of any change in behavior. Another timing aspect that should be noted is that when considers the experiment periods in weather and school-year contexts, we can assume that physical activity becomes more challenging as the experiment advances due to the North American winter conditions. In addition, for period 3 we can expect increased end-of-semester workload and stress for the student participants in the subject group.

Period	Dates	Days
1	Oct 5-Oct 27	1-23
2	Oct 28-Nov 15	24-42
3	Nov16-Dec 5	43-62

Table 2. Dates and days of periods used for analysis.

Normalized Activity Values

For analysis purposes, we normalized activity levels to the span of a single sample. For example, a normalized "daily average activity" is calculated by summing all accelerometer samples for the day and then dividing by the total count of accelerometer readings for the day. This gives us the average activity level per reading for that day. This allows us to easily do things like compare between normalized average activity levels in different times of day. It is trivial to convert a normalized value to actual time: For example, a normalized daily average value of 1.0 for an experimental group represents an average activity of 96 minutes per member.

Aggregated Activity Levels

One would reasonably assume that accelerometer readings would not be uniformly distributed throughout the day. A visual inspection of the distribution of non-zero readings indicated that that the day should be split the day into four quarters of six hours each, starting at midnight, in order to explore the difference in average accelerometer score per reading. Table 3 confirms that activity varies greatly throughout the day, and that it correlates with general intuition about the times of high and low activity.

Time of day	Average accelerometer score per reading
Midnight-6AM	0.23
6AM-Noon	1.29
Noon-6PM	2.34
6PM-Midnight	1.31

Table 3. Average accelerometer score by time of day. The average score per reading is much lower during the night and highest in the afternoon, as expected.

We refer to the a day's worth of accelerometer measurement for one person as a "person-day". For a single person, a complete day's worth of data was 720 accelerometer score readings, since accelerometer scans were taken in two-minute intervals. Data was considered "missing" for an interval if there was no accelerometer score logged for that interval. As also assumed in [16], we attribute most missing data to the phone being off, usually during night-time. As the current analysis deals primarily with daily average activity levels and change in daily average activity across time and experimental condition, we precluded person-days that did not have sufficient information for generating a reliable average score for the day. We observed that for days that had more than 50% of the possible readings, the missing datapoints were relatively uniformly spread accross the day, while in days with fewer than 50% of possible readings, they were not uniformally distributed and could not reliably be used. When a person's day had fewer than 50% of the possible readings, that day was not used for the analysis and calculation of averages. Removed measurements account for less than 5.4% of the available measurments.

Groups tested	Group 1 mean	Group 2 mean	D	p-value
Pre-Intervention (Period 1)				
Cntrl vs. PS&PR	1.162	1.241	0.3261	0.046*
Cntrl vs. PSee	1.162	1.266	0.3478	0.078
Cntrl vs. PRew	1.162	1.216	0.3043	0.164
PSee vs. PRew	1.266	1.216	0.2609	0.316
Post-Intervention (Periods 2 and 3)				
Cntrl vs. PS&PR	1.207	1.328	0.3718	0.001***
Cntrl vs. PSee	1.207	1.341	0.4193	0.001***
Cntrl vs. PRew	1.207	1.316	0.3590	0.007**
PSee vs. PRew	1.341	1.316	0.1026	0.976

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$

Table 4. Pairwise K-S comparison of activity level of the three experimental conditions pre- and post-intervention. The groups which are being compared are listed in the first column. "Group 1 mean" refers to the group listed first and "Group 2 mean" refers to the group listed second.

Activity Levels by Condition

Table 4 presents information about daily average activity levels in a pairwise comparison of the three experimental conditions pre- and post-intervention, using the K-S test. For this analysis, the two post-intervention periods are combined into one. Ideally, in the pre-intervention period we expect the null hypothesis to be true. While in the comparison that compares the control group vs. both social conditions the result is statistically significant ($p < 0.05$), in the direct pairwise comparison the test does not exhibit statistical significance, as expected. Conversely, according to the experimental mechanism design hypothesis, we anticipate that the social conditions will do better than the control, and possibly exhibit difference properties when compared to each other. For all comparisons between the social conditions (independently and jointly) and the control group, K-S test shows statistical significance ($p < 0.01$ and $p < 0.005$). However the difference between the two social conditions comes out non-significant under this comparison, possibly due to the inclusion of novelty effects through combining both post-intervention periods.

Reward efficiency

We are interested in the change in activity levels for each group rather than simple comparison of activity means. Furthermore, we want to evaluate the effectiveness of the exogenous money or energy injected into a system. We define "reward efficiency", η, which represents the activity change per dollar invested in the system. Reward efficiency for condition i is defined as:

$$\eta_i = \frac{\overline{a_{i,3}} - \overline{a_{i,1}}}{\overline{R_{i,3}}}$$

where $\overline{a_{i,k}}$ is the mean activity level for all participants in group i in period k, and $\overline{R_{i,k}}$ is the average reward per participant in group i in period k. Period 3 is used as the reference frame since we want to look at longer-term adherence. Tables 5 and 6 present information on reward efficiency for this dataset, based on actual monetary reward paid. Table 6 shows results of pairwise K-S testing of reward efficiency values, where all but one demonstrate statistical significance. In Table 5 we see that reward efficiency is more than doubled between the control condition and the Peer-See condition, and the efficiency of the Peer-Reward group is even more than the latter when comparing the conditions as a whole. In relative terms, we observe an average activity increase for Control, Peer-See, and Peer-Reward of 3.2%, 5.5%, and 10.4% respectively, counting in data from all times of day, days of week, sick-times, holidays, and so on. For the Peer-Reward condition, this comes down to an average increase of 84 minutes of physical activity per week, per participant.

As the underlying differences between the two social conditions were not clearly apparent, in Table 6 we dive into the social component. We divide the subjects according to their pre-reported closeness level with their Buddies. Although the overall comparison of the social conditions does not present statistical significance, the further grouping according to pre-existing relationships shows that the Peer-Reward condition achieves better results in two out of the three cases (close buddies and stranger buddies), while the Peer-See condition achieves better results for mixed buddies. For all these cases, we get statistical significance ($p < 0.01$). We see a complicated interaction element with regards to the Buddy closeness, which we touch on in the next section.

Condition	Activity Change from Period 1 to Period 3	Reward in Period 3	Reward Efficiency (Δ/\$)
Overall			
Control	0.037	\$3.00	0.012
Exp 1	0.070	\$2.77	0.0253
Exp 2	0.126	\$3.04	0.0416
Close Buddies (both Buddies score 3 or higher)			
Exp 1	0.118	\$2.68	0.0444
Exp 2	0.269	\$3.00	0.0896
Stranger Buddies (both Buddies score 2 or lower)			
Exp 1	-0.007	\$2.82	-0.0025
Exp 2	0.137	\$2.95	0.0464
Mixed Buddies (one Close, one Stranger)			
Exp 1	0.154	\$2.75	0.0560
Exp 2	0.053	\$3.12	0.0171

Table 5. Reward efficiencies (η). Reward efficiency is defined as the amount of activity level increase per dollar of reward paid.

Discussion

In this analysis we begin investigating the effectiveness of the different motivation and influence mechanisms for encouraging increased physical activity in-situ. We focus on two key metrics: The first is differences in average activity levels, both across conditions and chronological periods of the experiment, and the second is the efficiency of the reward "investment" in the system.

When daily average activity levels are analyzed, they support the hypothesis that the social components of both experimental conditions, together and separately, lead to a statistically significant positive difference. Analysis of the difference of effect between the two socially involved experimental groups is more complex, and dividing the experimental groups based on pre-intervention closeness of the Buddy triads reveals different trends. When reward efficiency is analyzed, we again see a significant difference between the control group on the one hand and the two experimental groups, taken together, on the other.

Results confirm our notions that embedding the social aspects in this non-competitive game adds to physical activity performance and activity adherence over time, compared to the socially isolated control condition An interesting question arises with respect to the social mechanisms. In the Peer-See group, there is social information that traverses the links between peered Buddies, but participants still receives a "selfish" reward. In Peer-Reward, both information and reward traverse the links between peers, and a potential for social influence as motivator. The intensity of pre-existing social relationships seems to play a factor, and results seem to support a complex contagion like phenomena, as described by Centola and Macy [10], especially when observing the interplay in triads where there is a "mix" of close and stranger peers. We have yet to investigate the communication patterns between the peers, and their subjective view of their condition, to try and understand if and how the social influence or pressure was exerted. We hope that by analyzing additional signals already collected, like the communication logs and co-location information, as well as related surveys administered to the participants, we will be able to shed more light on these underlying processes.

Had this intervention been conducted in springtime, one might expect a natural rise in physical activity as weather improves, which might have made it hard to separate the intervention's contribution. By going against the natural trend during winter, we challenge our experimental mechanisms. While results are not fully conclusive, they may suggest that while performance in the control and even Peer-See conditions deteriorates as time passes, the performance in Peer-Reward is slower to start but steadier in increase over time. The observations might support a hypothesis that the Peer-Reward

Groups being compared	Group 1 reward efficiency	Group 2 reward efficiency	D	p-value
Overall				
Cntrl vs. PSee	0.0120	0.0253	1.000	0.001**
Cntrl vs. PRew	0.0120	0.0416	1.000	0.001**
PSee vs. PRew	0.0253	0.0416	0.429	0.432
Close Buddies (both Buddies score 3 or higher)				
PSee vs. PRew	0.0444	0.0896	1.000	0.002**
Stranger Buddies (both Buddies score 2 or lower)				
PSee vs. PRew	-0.0025	0.0464	1.000	0.001**
Mixed Buddies (one Close, one Stranger)				
PSee vs. PRew	0.0560	0.0171	1.000	0.001**

** $p < 0.01$

Table 6. Testing significance in the differences between the reward efficiencies. All differences are statistically significant, except the difference between the two experimental groups when taken in their entirety.

condition induces social capital that takes time to build up, but once in place provides a more sustainable incentive structure than the direct monetary reward, or alternatively, a way to augment the exogenous monetary compensation with indigenous social capital, leading to a higher efficiency, and higher "return" on every Dollar invested in the system.

It is also important to mention that by design choice, we did not perform any external communication "scaffolding" to encourage social interaction. There were no mechanism within the study software for sharing results and promoting discussion - any such actions were done by participants on their own accord using their existing means of interaction. Related studies with social components [19, 11] suggest that adding explicit communication mechanisms to the technical system might add to the social effects of the intervention.

CONCLUSION

In this paper we introduced the Friends and Family Study, which combines high-dimensionality and high-throughput social and behavioral sensing using ubiquitous mobile phones, together with experimental interventions. We described our Android phone centred system that has been deployed in the study for over a year now. We presented initial results of a specific experimental intervention that demonstrates the great potential of the study dataset, its underlying technical system, and the of the general Social fMRI approach for measuring and experimenting with social mechanisms.

Through the fitness intervention example, we demonstrated challenges and benefits of leveraging our prior observations for the experiment design. We presented three key findings through this intervention: First, results support there is a statistically significant effect of social components on the real-world in-situ physical activity levels. Second, results show that our novel Peer-Reward social influence mechanism leveraging social capital can increase the efficiency of exogenous money and resources invested in the system. This could contribute to the design of future policies and intervention. Finally, we see a complex interaction effect related to pre-existing social ties inside the social experimental conditions. This could support hypothesis of a complex contagion like effect that should be further investigated. Immediate future work includes expanding the analysis of the existing data, as well as the design of new experiments based on these initial observations, particularly in the area of quantifying social capital and favor exchange. We hope that isolating and evaluating health related social mechanisms will become part of the toolbox for encouraging healthy behavior, combined with other components such as user interfaces, accurate measurement techniques, and individual goal setting.

In the same way that fMRI techniques help map the interconnections and mechanics of the human brain, we hope that our work will help advocate an evolution from mostly passive observatories to Social fMRI type of studies that can help further our understanding of the interconnections and mechanics of human society.

ACKNOWLEDGEMENTS

We would like to thank Sai Moturu, Cody Sumter, and Maya Orbach for their contribution towards the experimental deployment, surveys, and visualizations, and the undergraduate contributors for their help with the development of the sensing platform. We thank the Ubicomp reviewers for their thoughtful comments. Most importantly, we thank the Friends and Family Study participants.

REFERENCES

1. Cardiotrainer. http://www.worksmartlabs.com/cardiotrainer/about.php.
2. Funf: Open sensing framework. http://funf.media.mit.edu.
3. The human speechome project. http://www.media.mit.edu/cogmac/projects/hsp.html.
4. I. Anderson et al. Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. *Mob. Netw. Appl.*, 12:185–199, March 2007.
5. R. Barker and U. of Kansas. Midwest Psychological Field Station. *Ecological psychology: concepts and methods for studying the environment of human behavior*. Stanford University Press, 1968.
6. A. Bauman. Updating the evidence that physical activity is good for health: an epidemiological review 2000-2003. *Journal of Science and Medicine in Sport*, 7(1, Supplement 1):6 – 19, 2004. Physical Activity.
7. L. Berkman and T. Glass. Social integration, social support, and health. In L. Berkman and I. E. Kawachi, editors, *Social Epidemiology*. 2000.
8. S. N. Blair, Y. Cheng, and S. J. Holder. Is physical activity or physical fitness more important in defining health benefits? *Medicine & Science in Sports & Exercise*, 33(6), 2001.

9. C. V. Bouten, K. T. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Trans Biomed Eng*, Mar.

10. D. Centola and M. Macy. Complex contagions and the weakness of long ties. *The American Journal of Sociology*, 113(3):pp. 702–734, 2007.

11. S. Consolvo, K. Everitt, I. Smith, and J. A. Landay. Design requirements for technologies that encourage physical activity. In *Proceedings of CHI '06*, New York, NY, USA, 2006. ACM.

12. S. Consolvo et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceeding of CHI '08*, pages 1797–1806, New York, NY, USA, 2008. ACM.

13. T. Dawber. *The Framingham Study. The Epidemiology of Atherosclerotic Disease*. Harvard U. Press, 1980.

14. R. I. M. Dunbar. Co-evolution of neocortex size, group size and language in humans. *Behavioral and Brain Sciences*, 16(4):681–735, 1993.

15. N. Eagle, M. Macy, and R. Claxton. Network diversity and economic development. *Science*, 328(5981), 2010.

16. N. Eagle and A. Pentland. Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing*, (10):255–268, 2006.

17. N. Eagle, A. S. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *Proc. Natl. Academy of Sciences*, 106(36), 2009.

18. R. G. Eston, A. V. Rowlands, and D. K. Ingledew. Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children's activities. *Journal of Applied Physiology*, 84(1):362–371, 1998.

19. D. Foster, C. Linehan, and S. Lawson. Motivating physical activity at work: using persuasive social media extensions for simple mobile devices. *Design*, 2010.

20. J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong. Mylifebits: fulfilling the memex vision. MULTIMEDIA '02, New York, NY, USA, 2002.

21. M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 06 2008.

22. C. A. Hidalgo and C. Rodriguez-Sickert. The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications*, 387(12), 2008.

23. O. P. John and S. Srivastava. *The Big-Five trait taxonomy: History, measurement, and theoretical perspectives*, volume 2, pages 102–138. Guilford, 1999.

24. D. Lazer et al. Computational social science. *Science*, 323(5915):721–723, 2009.

25. J. J. Lin, L. Mamykina, S. Lindtner, G. Delajoux, and H. B. Strub. Fish'n'steps: Encouraging physical activity with an interactive computer game. In *Proc. of the International Conf. on Ubiquitous Computing*, 2006.

26. H. Lu et al. The jigsaw continuous sensing engine for mobile phone applications. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, SenSys '10, New York, NY, USA, 2010.

27. A. Madan, M. Cebrian, D. Lazer, and A. Pentland. Social sensing for epidemiological behavior change. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 291–300, New York, NY, USA, 2010. ACM.

28. A. Madan, K. Farrahi, D. G. Perez, and A. Pentland. Pervasive sensing to model political opinions in face-to-face networks. In *Pervasive'11 (in press)*, 2011.

29. J. Maitland and K. A. Siek. Technological approaches to promoting physical activity. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction*, New York, 2009. ACM.

30. R. Montoliu and D. Gatica-Perez. Discovering human places of interest from multimodal mobile phone data. In *Proc of 9th Int. Conference on on Mobile and Ubiquitous Multimedia (MUM,',',')*, 12 2010.

31. D. O. Olguín, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 39(1):43–55, 2009.

32. J.-P. Onnela and F. Reed-Tsochas. Spontaneous emergence of social influence in online systems. *Proc. Natl. Academy of Sciences*, 107(43), 2010.

33. S. T. Saponas et al. ilearn on the iphone: Real-time human activity classification on commodity mobile phones. Tech report, University of Washington, 2008.

34. T. Toscos et al. Encouraging physical activity in teens can technology help reduce barriers to physical activity in adolescent girls? In *Pervasive Computing Technologies for Healthcare*, 2008.

35. R. P. Troiano et al. Physical activity in the united states measured by accelerometer. *Med Sci Sports Exerc*, 40(1):181–8, Jan 2008.