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 Laboratory, 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 [11], where we discovered that the causality may go along the opposite direction than proposed. 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 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 [8] 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 [23]. describe the potential of computational social science to increase our knowledge of individuals, groups, and societies, with an unprecedented breadth, depth, and scale. [23] 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 [9], which stands out for its duration and a subject pool of several thousands, however its "throughput" is very 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 [17]. Eagle et. al find that the diversity of individuals relationships is strongly correlated with the economic development of communities [11]. 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, [19], 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[29]. 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 [12] 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 [12, 13]. This initial study was then expanded in Madan et al.[25], 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 [26]. Other works using mobile phones as sensor platforms were done by Montoliu, et. al [27], Lu, et. al [24], and from CENS center, e.g. Campaignr by Joki et. al [22]. These have yet to take the further step of a living lab integrated with tailored experimental interventions. 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 [28]. An important aspect of the Sociable Badges is that they have also been deployed in studies where sensor feedback was given to the corporate participants [28].

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., and proposed explanations link the problem to societal changes that promote less physical ac-
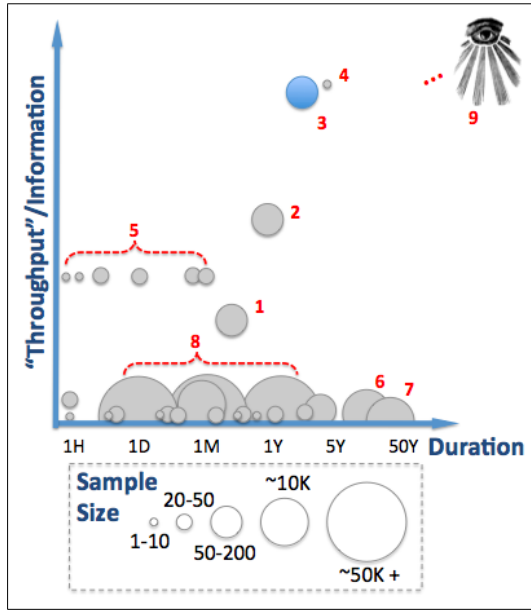
Figure 1. Qualitative Overview of social science "observatories" and datasets, along axes of data collection duration, qualitative "throughput" , and the size of the subject sample.(1)Reality Mining[12], (2) Social Evolution[26], (3) Friends and Family Dataset, (4) rich-data pioneers [2, 16] (5) Sociometric Badge studies [28], (6) Midwest Field Station [3], (7) Framingham Heart Study[9], (8) Large Call Record Datasets [17, 11, 7, 19] , (9) "Omniscient"/All-Seeing View

tivity and increased food consumption [18, 20], Physical activity has been shown to help alleviate the burden of obesity and other health conditions [4, 5]. However, there is still a lot to learn 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.

Over the past two decades, the accelerometer has been established and refined as a tool for tracking physical activity [6, 14, 30]. 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 [6]. In fact, accelerometer data has been found to provide more accurate estimates than other widely-used proxies for energy expenditure [14], and it has become accepted practice to use accelerometer data as a reliable proxy for energy expenditure in academic studies [30].

Nowadays, 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. Finally, some works have begun exploring the sharing of physical activity information via social networking tools, like Foster et. al's Facebook based workplace experiment [15].

THE SOCIAL ƒMRI

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 [12], adding throughput and dimensionality while incorporating a strong element of active interaction and carefully directed 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 [12], and undergraduates in [25]. 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.

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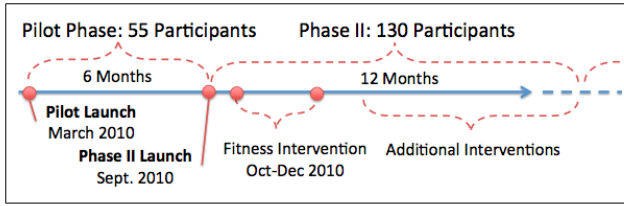
Figure 2. High level timeline for the Friends and Family study.

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

well with Dunbar's social evolutionary theory regarding the number of people humans are able to maintain a relationship with [10]. 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 [21]. 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.
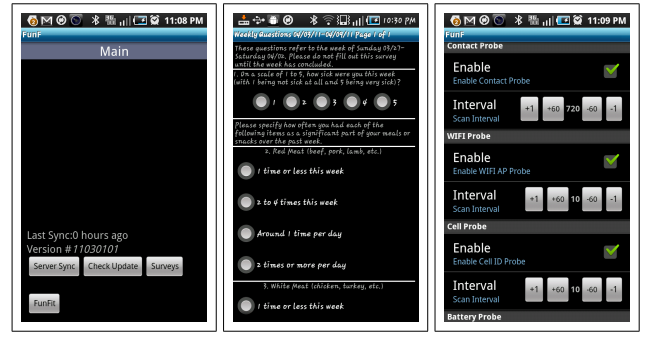
Facebook Data Collection Application

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. There are currently over 25 different types of data signals collected by the phone platform. The system also supports integration of user-level apps, such as 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. '

Data Formats and Server Communications
Phone data is is saved in SQLite file format. Files are rotated every three hours to prevent large loss due to file corruption and to allow periodic transmission of the data to the

4

study's central server, located in a secure physical location. 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 (for example a participant connect to a WiFi access point to browse the web), the system attempts to upload the files. Once uploaded, files are also encrypted and saved in a backup directory on the phone, until data integrity is confirmed on the server side. In addition to uploading file from the server, the system downloads from the server parameter configurations files, which 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.

Battery Considerations

The probe configuration is set so that battery-intensive actions (e.g. GPS scans) are performed in intervals allowing usefulness while minimizing battery drain. Remote configuration capability allows for fine-tuning the system, and our goal was to enable a minimum of 16 hours between charges.

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

In the control condition, subjects saw their own progress as visualized in Figure 4. Reward given to the subjects depended only on their own activity.
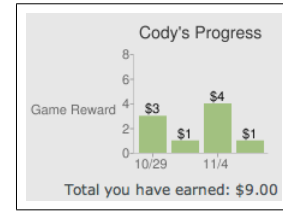

Figure 4. Reward display for participants in the Control condition

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 the progress of two Buddies, but this time subjects' rewards actually 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 for C. The Peer-Reward feedback page displays how much reward they are getting from their Buddies, as well how much reward they are earning for the people that they are Buddy to.

Accelerometer-based Activity Measurement

The key signal used in this intervention is the phone's 3-axis accelerometer. We care mostly about relative change in activity based on different conditions, and are not attempting to get accurate pedometer and activity classification. Therefore, we decided to implement a less accurate but more robust algorithm for estimating physical activity levels. This allowed for increased flexibility in the way participants can carry the phone, and reduces 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 [14]. The score was calculated by giving one point for every second thresholded to two states 1)"still" and 2)"moving". The score was calculated by giving one point for every second when the variance of the magnitude of acceleration was above a certain threshold.

Game Reward Calculation

To calculate the reward for the B-day bin ending in day T, where B is bin size, we calculate the person's average accelerometer measurement over the bin days: $\bar{a}_{[T-B,T]}$. We also define a sliding window of size W days that ends with

Figure 5. Reward Function

day $T - B$ which is used for calculating the baseline we compare against. In our case, $B = 3$, and $W = 7$. We estimate each person's average activity μ_{w_T} and variance σ_{w_T} from the time window $a_t, \forall t \in [T - B - W, T - B - 1]$. The activity range between $[\mu_{w_T} \pm \sigma_{w_T}]$ was divided into 10 equal intervals, and each was mapped to values between \$0.50-\$5, in \$0.50 increments, according to the function defined in Figure 5. 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. However, investigators have to take measures that will not interfere with people's daily routine or lead to any human-subjects risks. The experiment has 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 performance), 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 and has motivation 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 strong friendship ties (closeness > 2) in the study pool. The mean distance between any two participant is 1.9.

Experimental Condition Allocation
Based on this information and the information on couples in the study, we designed an allocation algorithm: For participants $p_1, \ldots, p_N$ in either the Peer-See or the Peer-Reward group, we match each participant p_i with two buddies. p_i will monitor the progress, and in the Peer-Reward case, get paid for the progress of her buddies. We set $b_j^i = 1$ if p_j is p_i's buddy, 0 otherwise. Due to symmetry, each participant also naturally has two other participants to monitor her.

The participant self-reported closeness is captured in an adjacency matrix $\mathbf{R}$, where $\mathbf{R}_{i,j}$ captures the closeness between p_i and p_j. Another matrix $\mathbf{M}$ captures the marriage status between two participants, where $\mathbf{M}_{i,j} = 1$ if p_i and p_j are married/partners, 0 if not. Our goal is to allocate closest friends for each participant as buddies, while keeping the designed network topology. This can be written as a binary integer programming formulation:

$$\arg \max_{b_j^i, i, j \in \{1, \ldots, N\}} \sum_{i=1,\ldots,N} \sum_{j=1,\ldots,N} b_j^i \mathbf{R}_{i,j} \qquad (1)$$

$$\text{subject to: } \forall i, \sum_{j=1,\ldots,N} b_j^i = 2,$$

$$\forall i \in \{1, \ldots, N\}, b_i^i = 0,$$

$$\forall i, j \in \{1, \ldots, N\}, b_j^i + b_i^j < 2,$$

$$\sum_{i=1,\ldots,N} \sum_{j=1,\ldots,N} \mathbf{M}_{i,j} b_j^i = 0.$$

The first constraint forces each participant to have exactly two buddies. The second constraint is set so that participants can not be buddy of themselves. The third constrain prohibits two participants from being buddies of each other. The last constraint means that participants cannot take their spouses as buddies. This decision eliminates the unique and complicated effects resulting from marriage ties, and it ensures that our fitness peer monitoring topology is purely constructed on regular friendship ties.

The integer programming problem cannot be solved directly, and we apply an iterative approach: In each iteration, we randomize initial values and use 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 dropped out and were removed from the study pool over the course of the intervention. 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 (October 5 through October 27), the first 19 days of the intervention (October 28 through November 15), and the second 20 days of the intervention (November 16 through December 5). The periods are summarized in Table 2. The days after the intervention begins are broken up into two periods in order 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.

Daily Average Activity Levels

The main accelerometer metric that we used is the a "daily average activity" measure, $\overline{a}$. It is calculated by summing all accelerometer samples for the day and then dividing by the total count of accelerometer readings for the day. The daily average activity level can thus be understood as the average activity level per reading for a given day. Accelerometer measurements are performed continuously and recorded to the database every two minutes for a total of 720 possible readings per day per individual.

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 [12], 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 dayes 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..

Activity Levels by Condition

Tables 4, 5 and 6 present information about daily average activity levels. Table 4 shows 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.

In Table 5 and 6 we investigate change in mean activity level between the pre-intervention period and the second intervention period for the Control and the two experimental conditions, respectively. As the underlying differences between

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.

	Control	
Period	Activity	Δ
Overall		
1	1.162	—
3	1.199	+0.037

Table 5. Mean activity level and difference, Δ, for periods 1 and 3 for Control condition. K-S test not significant.

	Peer-See		**Peer-Reward**	
Period	Activity	Δ	Activity	Δ
Overall				
1	1.266	—	1.216	—
3	1.336	+0.070*	1.342	+0.126
Close Buddies (both Buddies score 3 or higher)				
1	1.545	—	0.966	—
3	1.663	+0.138	1.235	+0.269**
Stranger Buddies (both Buddies score 2 or lower)				
1	1.349	—	1.238	—
3	1.342	-0.007	1.375	+0.137*
Mixed Buddies (one Close, one Stranger)				
1	1.021	—	1.321	—
3	1.175	+0.154*	1.374	0.108

$* \, p < 0.05 \,; ** \, p < 0.01$

Table 6. Mean activity level and difference, Δ, for periods 1 and 3 for for the two experimental conditions, overall and divided by closeness of Buddy triads. Tested for significance using K-S test.

the two social conditions were not clearly apparent, we dive into the social component. We divide the subjects according to their pre-reported closeness level with their Buddies. The most interesting observation is that the preexisting social relationship between for both social conditions are not consistent. We touch on this in the "Discussion" section.

Reward efficiency
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 7 and 8 present information on reward efficiency for this dataset, based on actual monetary reward paid. Table 8 shows results of pairwise K-S testing of reward efficiency values, where all but one demonstrate statistical significance. In Table 7 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. Although that last result does not present statistical signifi-

cance, the further grouping according to pre-existing social 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 ($p < 0.01$). Again we see a complicated interaction element with regards to the Buddy closeness.

Discussion
In this analysis we start 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. The differences in reward efficiency is more apparent than the differences in activity level.

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

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

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 8. 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.

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 [8], 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 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. The Facebook fitness study results of [15] 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.

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. 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.

REFERENCES

1. Cardiotrainer.
   http://www.worksmartlabs.com/cardiotrainer/about.php.

2. The human speechome project.
   http://www.media.mit.edu/cogmac/projects/hsp.html.

3. 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.

4. 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.

5. S. N. BLAIR, Y. CHENG, and J. SCOTT HOLDER. Is physical activity or physical fitness more important in defining health benefits? *Medicine & Science in Sports & Exercise*, 33(6), 2001.

6. 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.

7. M. Cebrian, M. Lahiri, N. Oliver, and A. Pentland. Measuring the collective potential of populations from dynamic social interaction data. *Selected Topics in Signal Processing, IEEE Journal of*, 4(4), 2010.

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

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

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

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

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

13. 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.

14. 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.

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

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

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

18. A. A. Hedley, C. L. Ogden, C. L. Johnson, M. D. Carroll, L. R. Curtin, and K. M. Flegal. Prevalence of overweight and obesity among us children, adolescents, and adults, 1999-2002. *JAMA: The Journal of the American Medical Association*, 291(23), 2004.

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

20. J. O. Hill and J. C. Peters. Environmental contributions to the obesity epidemic. *Science*, 280(5368), 1998.

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

22. A. Joki, J. A. Burke, and D. Estrin. Campaignr: A framework for participatory data collection on mobile phones, 2007.

23. D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. Computational social science. *Science*, 323(5915):721–723, 2009.

24. H. Lu, J. Yang, Z. Liu, N. D. Lane, T. Choudhury, and A. T. Campbell. 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.

25. 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.

26. 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.

27. 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.

28. 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.

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

30. R. P. Troiano, D. Berrigan, K. W. Dodd, L. C. Mâsse, T. Tilert, and M. McDowell. Physical activity in the united states measured by accelerometer. *Med Sci Sports Exerc*, 40(1):181–8, Jan 2008.