Measuring and Generating Social Influence Using Mobile Telephony

by

CORY M. IP

B.E., 2008
The City College of New York

A.B., 2004
Harvard University

Submitted to the Engineering Systems Division
in partial fulfillment of the requirements for the degree of

Master of Science in Technology and Policy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2011

Author..
Engineering Systems Division
January 21, 2011

Certified by..
Alex (Sandy) Pentland
Toshiba Professor of Media Arts & Science
Thesis Supervisor

Accepted by..
Dava J. Newman
Professor of Aeronautics and Astronautics and Engineering Systems
Director, Technology and Policy Program

Measuring and Generating Social Influence Using Mobile Telephony

by

CORY M. IP

Submitted to the Engineering Systems Division
on January 21, 2011, in partial fulfillment of the
requirements for the degree of
Master of Science in Technology and Policy

Abstract

In recent years, mobile telephony and manufactured social influence have received attention as tools for eliciting behavior change. This paper describes an experimental approach to studying the effect of these tools in combination. We demonstrate that mobile telephony can be used to generate social influence and elicit behavior change in support of a specified goal. A longitudinal experiment is conducted in a free-living environment, using mobile telephony to track and attempt to influence the activity level of a pool of 97 subjects over 62 days. Subjects receiving feedback about others' performance show a significantly greater increase in activity level than subjects receiving feedback about their own performance only, demonstrating that social influence enhances the persuasive capability of mobile telephony. A significant effect is observed of a subject's pre-existing closeness to the others whose performance he sees. Evidence is also seen that generated social ties lead to a more significant and sustained increase in activity level than existing social ties.

Thesis Supervisor: Alex (Sandy) Pentland
Title: Toshiba Professor of Media Arts & Science

Acknowledgments

To my adviser, Sandy: Thank you for the opportunity to join in your research effort, if only for a little while. Working with you has opened up entirely new areas of interest and experience for me, and I still think your research is as fascinating as the day I first hunted you down in your office. Thank you for your guidance during the research and thesis process, your willingness to meet anytime, and especially your speedy responses when it got down to crunch time.

To my research group, especially Nadav and Wei: Thank you for your patience and partnership over the past year. It has been a privilege, not to mention great fun, to work with a bunch of hackers.

To the TPP leadership and administration, including Dava Newman, Frank Field, Sydney Miller, and Ed Ballo: Thank you for all the effort you put into making TPP such a wonderful program. A special thank you to Sydney, for being a shoulder to lean on during my second-year mid-life crisis.

Mommy and Daddy: Thank you for being proud of me no matter what, and for making me want to earn it.

Boma: Your advice and support through this process has been invaluable. Thank you not only for your wisdom about the thesis process itself, but for always being reassuring, no matter how dire my straits were. And of course, thank you for every inspirational song and YouTube video you sent my way.

My aMAAzing ladies, Allie, Ellie, and Ruth: I am so thankful we ended up in TPP together. Thank you for all the dinner parties, for all the clothes we shared, and for every awesome group birthday present. We'll always have Niagara!

Last and most important, Garth: Thank you for everything. This thesis wouldn't exist without you, in more ways than I care to acknowledge. Beyond that, thank you for every time you pushed me to think harder to understand an idea or defend a thought, even when I acted like it was torture. Over the past two and a half years, I have learned so much from you about who I am and who I want to be. I do not have the words to tell you how grateful I am to have you in my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In many of life's pursuits, across ages and professions, success hinges on being able to elicit changes in people's behavior. Behavior change can be sought for goals more or less noble, from the policymaker trying to encourage responsible savings decisions to the marketer trying to popularize the latest line of shampoo. In recent years, two different tools for eliciting behavior change have been the objects of a surge in academic and corporate interest: manufactured social influence and mobile telephony.

1.1 Manufactured social influence

Humans are not completely rational creatures in all of their choices and behaviors. This characteristic creates the possibility of influencing human behavior in ways both overt and subtle. Marketers have long taken advantage of the psychological weaknesses of humans. Free trials, as an example, exploit the human tendency to form an attachment to an object as soon as one possesses it. Social psychologists have engaged in more rigorous study of what motivates behavior, with recent decades seeing a marked increase in experiments demonstrating a link between social norms and behavior. Legal scholars have also begun to address how social norms and influence bear on legal issues, recognizing the fact that it is easiest to achieve compliance when laws are aligned with norms (Mullainathan, Schwartzstein, and Shleifer 2008). Taking that line of thinking even further, some legal scholars have argued that we should try to achieve target behaviors by activating norms toward that behavior, without having to institute laws (Thaler and Sunstein 2008). We refer to this as manufactured social influence, rather than simply social influence, to distinguish between social influence generally, which is always present, and social influence which is purposefully activated or generated in order to elicit a specified behavior.

1.2 Mobile telephony

In the past decade, the potential of mobile telephony has caught the attention of professionals of all stripes, from businessmen to health professionals to academics. The applications of mobile telephony in business are straightforward. Mobile phone apps that harness the ubiquity and portability of mobile phones can be sold for profit. Health care professionals recognize the usefulness of mobile phones as a tool to encourage compliance with certain health behaviors, again because of its ubiquity and portability. In recent years, academics have also begun to recognize the power of mobile technologies more broadly and mobile phones specifically as tools for studying human behavior (Eagle and Pentland 2006; Sohn et al. 2006; Miluzzo et al. 2008). For almost a century, social scientists have relied on tools such as surveys and field studies to learn about human behavior, which had certain known limitations. Automatically collected data from mobile phones offers dense, continuous data of a kind social scientist have not previously had access to, and a growing number of researchers believe that this access to previously unimaginable and comprehensive datasets will revolutionize the study of social science, as it has biology or physics (Lazer et al. 2009). Studies have been designed that look not only at mobile technology from the perspective of the individual (e.g. what design features people respond best to in mobile applications, or under which circumstances people choose to share location data), but also from the perspective of what we can learn about human interaction. These data allow for fresh and deeper analysis of issues typically tackled by social scientists.

1.3 Thesis questions

The questions this thesis explore build on the two characteristics of mobile phones mentioned above: its potential as a tool for social science and its potential as a tool for influencing behavior, and especially how these two potentials combine.

1. **Question 1:** How do we best use the mobile phone as a social science tool? Can we not only corroborrate but enhance existing findings?

2. **Question 2:** How do we best use the mobile phone as a tool for persuasion and influencing behavior? How do we design a free-standing, low-effort application that will give us the most bang for our buck?

   Additionally, we realized that mobile phones could not only track human interaction and social influence, but it could itself be a generator of social influence, leading to our third question:

3. **Question 3:** How can we use the mobile phone specifically to generate social influence, and will that enhance its usefulness as a tool for influencing behavior?

We decided to explore these questions through a longitudinal human subjects experiment, conducted in a free-living (as opposed to laboratory) setting. Our innovation is to combine the two persuasive techniques of social influence and mobile technologies, and Chapter 3 will elaborate on the details of the experiment.

1.4 Thesis Roadmap

The next chapter will set the stage for the questions we are exploring by giving an overview of the two major areas of exploration which these questions have been drawn from: first, the study of social influence and social norms in determining human behavior; and second, developments in persuasive technology. Chapter 3 provides details of the experiment we designed in order to test our hypotheses about these questions, and Chapter 4 presents the results of the experiment. Chapter 5 revisits the original questions in light of the data we've gathered and addresses implications for those wishing to leverage mobile technology for behavior change, and Chapter 6 discusses the contributions of the work, and future areas to explore based on the work done here.

Chapter 2

A Review of Behavior Change

At its most essential, this thesis aims to address the question of how we can change human behavior. We specifically look at the potential of combining the tools of social influence and mobile phones. However, this is one technique in a broader mission, which is to understand human behavior in an effort to help people live the best lives they can.

2.1 The study of human behavior

Human behavior is complex, and when broken down even the smallest action seems miraculous. The simplest conception of human behavior conceives of an intention translating directly into an action, but no one really thinks the process is that straightforward. Instead, human behavior can be broken down into several stages (Gollwitzer and Bargh 1996). First, the sources of our goals and intentions have to be investigated. Furthermore, there are affective influences on our goals and intentions – that is, feelings and emotions are often strong influences on our goals. Then we have to turn our intentions into action – a step that is by no means guaranteed. In light of all this, predicting and understanding human behavior seems a daunting if not impossible task.

However, the ability to predict and on top of that to understand human behavior is critical for the success of most human enterprise. The smooth functioning of society depends on compliance with government directives. Anyone who has marveled at the extraordinary coordination of a string of cars pulling over to the side of the road to allow an ambulance to pass has appreciated how compliance with norms contributes to an orderly society. From energy conservation to responsible financial planning to voter turnout, many of the most important policy questions our elected representatives face require solutions that must create a change in human behavior.

2.1.1 Predicting human behavior

There are two levels of influencing human behavior: first, we can be able to predict behaviors, and second, we can understand what causes behaviors. There are a number of professions whose practitioners depend on being able to influence human behavior for their livelihood. Cialdini and Trost (1998) calls these people "commercial compliance professionals," defined as anyone whose business or finanical well-being depends on their ability to induce compliaince, where compliance means acquiescence to a request. Examples of such professionals are salespeople, fund-raisers, advertisers, political lobbyists, negotiators, or con artists. Looking at the practices of these compliance professionals tells us that there are clear patterns in how people behave and how they respond to certain prompts. Cialdini and Trost (1998) identifies six tendencies that commercial compliance professionals routinely exploit in order to elicit desired behavior:

1) the tendency to reciprocate a gift, favor, or service;
2) the tendency to be consistent with prior commitments;
3) the tendency to follow the lead of similar others;
4) the tendency to accommodate the requests of those we know and like;
5) the tendency to conform to the directives of legitimate authority; and
6) the tendency to seize opportunities that are scarce or dwindling in availability.

The actions of these compliance professionals fall in the cateogry of predicting rather than understanding human behavior.

In some cases, policymakers may find that mere prediction of human behaivor suffices to solve their problem. In most cases, though, and especially if policymakers want to produce change, they must go beyond prediction to an attempt to understand human behavior.

2.1.2 Social psychologists: frameworks for understanding human behavior

The study of individual to group-scale human behavior has also been studied more academically. Over the years, a number of social psychology theories have emerged to propose a framework within which to understand human behavior, especially reflective behavior.[1] Here we give an overview of one of the most influential frameworks to understand such reflective behavior.

[1] Here, reflective behavior is in contrast with impulsive behavior. There is general agreement that reflective processes and impulsive processes are cognitively distinct, but there is considerably less agreement about how these two processes combine in producing behavior (Strack and Deutsch 2004). Because this thesis deals centrally with reflective, goal-oriented behaviors, we do not dive into the literature on impulsive cognitive processes, though this should be viewed as determined by space restrictions rather than any judgment of importance. Indeed, less conscious influences on our decision processes (right

2.1.2.1 Fishbein and Ajzen - The Theory of Reasoned Action, 1975

One very influential theory has been Fishbein and Ajzen's Theory of Reasoned Action(Fishbein and Ajzen 1975; Ajzen and Fishbein 1980). Fishbein and Ajzen had a bold goal: to put forth a single theoretical framework to explain "virtually any human behavior" (Ajzen and Fishbein 1980). This approach was unifying in social psychology at the time, for people had previously assumed that different behaviors had different causes. Their approach is elegant in limiting the theory to a few key concepts and relationships, as shown in Figure 2-1.

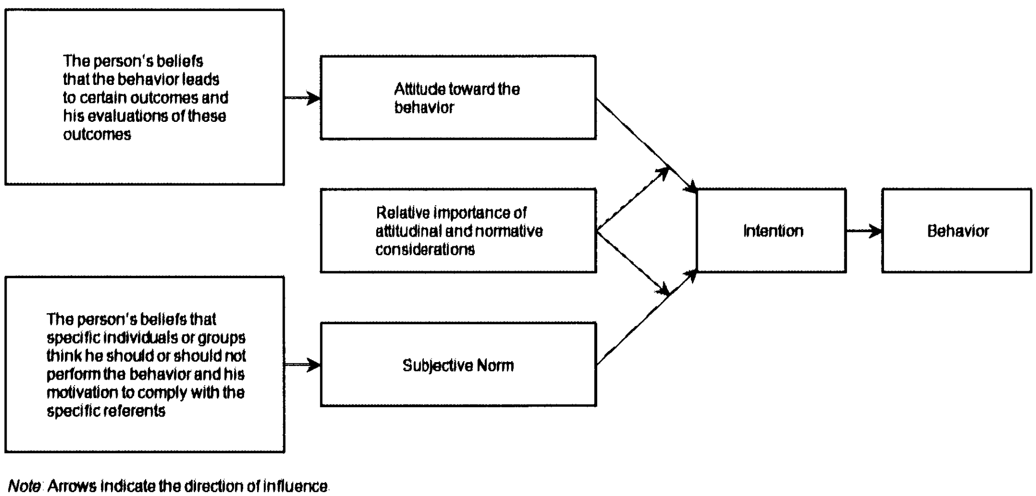

Figure 2-1: Schematic of Fishbein and Ajzen's Theory of Reasoned Action.
Source: Fishbein and Ajzen, 1975

Their theory successfully unites three core areas of study in social psychology: attitudes, social norms, and behavior ("action," in the name of the theory). It also offered several insights and new understandings which gave people a new conceptual framework within which to understand how these three siginificant areas combine to create action.

First, Fishbein and Ajzen believed that intention is strongly correlated with behavior. Taking for granted a strong correlation between intention and behavior. The mystery, then, lies in what forms the intention.

Second, TRA was revolutionary in its incorporation of the idea that it is attitude toward the outcome of a behavior, rather than attitude toward the target of the be-

now kept distinct from impulsive cognitive processes, but of the same spirit) are at the core of the paper.

havior, that determines intention. Take the example behavior of taking medicine for my cold. Even if I do not like medicine (the target of a behavior), I might still take the medicine because I believe that getting better (the outcome of the behavior) is good. This idea was not itself new. The first study establishing this disconnect between attitude and behavior was Richard LaPiere's 1934 investigation of racial prejudice. In this study, LaPiere surveyed 251 service establishments (restaurants, hotels, etc.) around the United States, asking them the question: "Will you accept members of the Chinese race as guests in your establishment?" Ninety percent (90%) of the 128 respondents answered no. The twist was that, six months earlier, LaPiere had visited all 251 establishments with a young Chinese couple, and they had been refused service only once. Fishbein and Ajzen incorporated this understanding for the first time into a theory for understanding and predicting human action.

A third key contribution of this theory is the inclusion of normative influence. This element has been the subject of controversy, as some parties questioned whether there was really a distinction between normative attitude and behavioral attitude (Miniard and Cohen 1981; Liska 1984), while studies which applied this theory could not establish as strong a correlation (if any) between normative influence and intention as there was between behavioral attitude and intention (Farley et al. 1981; Ajzen 1991).

Much of the problem stemmed from the recommended framework for establishing beliefs, subjective norms, and intentions. Fishbein and Ajzen developed an extended survey methodology to go along with the framework (Ajzen and Fishbein 1980). This was, however, more a limitation of the field at the time at which they were studying than an indictment of the theory itself. Despite the lack of quantitative evidence, a strong intuition that normative influence should be a predictor of intention and thus behavior persisted, and people did not dismiss norms as a determinant of behavior despite the lack of evidence proving the relationship (Terry and Hogg 1996; Trafimow and Finlay 1996; Finlay et al. 1997).

2.1.2.2 Bandura's concept of self-efficacy and Socio-Cognitive Theory, 1982

Bandura did not seek to give as comprehensive an explanation, but rather to highlight a factor that was perhaps neglected. He was addressing the same issue we've noted above, that people often know what the optimal behavior would be but do not execute it. A classic example is the case of exercising. Many people believe that exercising more would be optimal, but do not act on that belief.

One of Bandura's central concepts is that of self-efficacy (Bandura 1982). This concept captures how people's ability to think mediates the relationship between knowledge and action. Self-efficacy percepts are people's beliefs about their capabilities to exercise control over their own level of functioning and over events that affect their lives. When people have strong self-efficacy percepts, they are more likely to turn knowledge and intention into action. Conversely, weak self-efficacy percepts reflect a belief that one cannot one's environment, leading to a weaker tie between intention and action.

2.1.2.3 Ajzen's Theory of Planned Behavior, 1991

The Theory of Planned Behavior is essentially the Theory of Reasoned Action modified to include the concept of self-efficacy. It is modified and termed "perceived behavioral control," and more specifically addresses people's beliefs not about control over their lives in general but rather over their ability to perform a particular behavior.

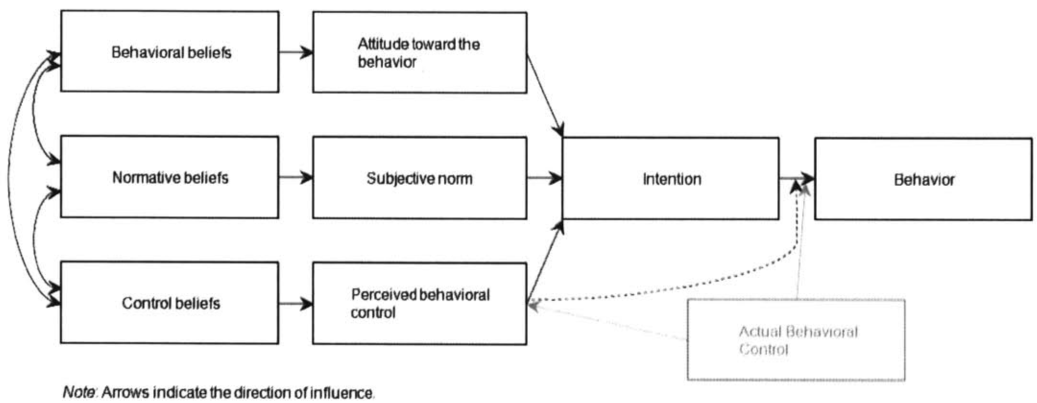

Figure 2-2: Schematic of Ajzen's Theory of Planned Behavior.
Source: Ajzen, 1991

In this updated schematic, perceived behavioral control and actual behavioral control modify the link between intention and behavior. This additional feature can help explain why intention and behavior do not correspond perfectly to attitudes and subjective norms, but does not address the weakness of quantitative evidence correlating subjective norm to behavior.

2.1.2.4 Summary of theories

Throughout these efforts, a consistent feature has been the difficulty of incorporating and then properly quantifying social influence effects. Though these different frameworks have acknowledged the importance of social influence on behavior, they have not been able to explain it satisfactorily. A component of the difficulty has been access to the right kind of data to study these questions. This relates directly to our first central question of the role that mobile phones can play in collecting data to answer social science questions.

2.2 Social norms as leverage for behavior change

Despite the fact that studies were not able to quantify a strong relationship between normative influence and behavior within the TRA survey framework, by the end of the twentieth century, there was growing interest in exploiting patterns of human behavior and responses to social influence, as compliance professionals had been doing for decades. A slew of studies looked at the effects of norms in field experimental settings rather than by relying on survey data, as Ajzen had advised, with more positive results. By the beginning of the 2000s, it was accepted that social norms guide action in direct and meaningful ways (Schultz et al. 2007).

2.2.1 Responses to normative influence

The past few decades have seen a boom in studying how we can use social influence to achieve goals. An avalanche of studies has proven that norms matter and can have significant effects in terms of behavior change. Particularly, we want to focus on the idea that seeing what other people do matters, whether we are aware of it or not.

One of the earliest such studies was a field study on littering conducted by Cialdini, Reno, and Kallgren (1990). The study led to two important clarifying principles about how norms act on behavior. First, Cialdini et al. (1990) identified two separate types of norms that could possibly work against each other, depending on the situation. Second, the subject's attention had to be focused on the norm in order to have an impact.

The two kinds of norms that Cialdini et al. (1990) identified were descriptive and injunctive norms. Descriptive norms are norms about what other people do. These contrast with injunctive norms, which are beliefs people hold about what others would approve or disapprove of. Put more succinctly, descriptive norms are beliefs about what is done, while injunctive norms are beliefs about what ought to be done. In this study, Cialdini et al. looked at littering behavior. Experimenters placed fliers on windshields of cars in a parking structure. In order to activate certain descriptive norms, two environments were set up: a clean environment and a littered environment. On top of this, to activate norm salience, experimenters had two different social contact conditions for people (unwitting subjects) walking to their cars. In the first condition, a third party was sent to merely walk past the subject. In the second condition, the third party was reading a flier identical to the one placed on the subject's windshield, and the subject witnessed the third party throwing the handbill on the ground. (An observer judged whether the subject had noticed the littering incident.) Thus, there were a total of four conditions:

1) anti-littering descriptive norm (clean environment), low norm salience (third party walks by)

2) anti-littering descriptive norm (clean environment), high norm salience (third party litters)
3) pro-littering descriptive norm (littered environment), low norm salience (third party walks by)
4) pro-littering descriptive norm (littered environment), high norm salience (third party litters)

The percentages of subjects littering in each of the four conditions is presented in Figure 2-3.

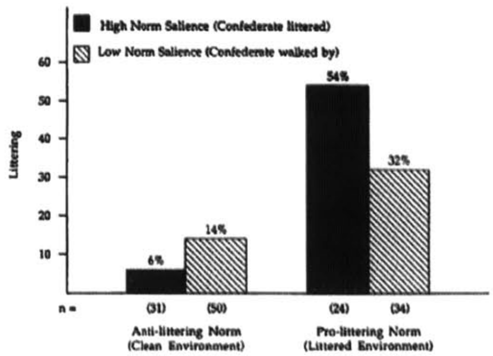

Figure 2-3: Percentage of subjects littering in each condition.
Source: Cialdini et al., 1990

The numbers reinforce the idea that the salience of a norm has significant influence on action. When the descriptive norm dictates that we should not do something, and our attention is drawn to it (clean environment, third party litters condition), we are LEAST likely to engage in that activity. Conversely, when the descriptive norm dictates that it is okay to do something (even if it is at odds with our injunctive norm), and our attention is drawn to that descriptive norm, then we are MOST likely to engage in that activity.

The injunctive norm concept that Cialdini introduced was tested in a different study about energy consumption behavior (Schultz et al. 2007). In this study, energy consumers were presented information about their energy usage compared to the average neighborhood usage. Researchers were surprised to find a "boomerang effect" when people were told that their energy consumption was below average. In other words, a descriptive norm that it was okay to consume more than they were consuming was made more salient, and they responded by increasing their consumption. The researchers eliminated this boomerang effect by combining the descriptive with an injunctive norm that they ought to continue their below-average consumption. This injunctive norm was activated by the inclusion of a smiley face. Above-average consumers in the injunctive condition were shown a picture of a frowning face, as shown in Figure 2-4.

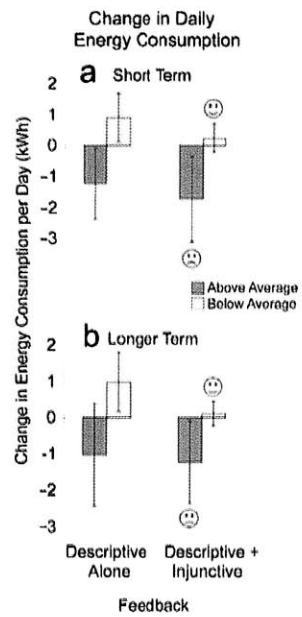

Figure 2-4: Change in daily energy consumption with different types of norms.
Source: Schultz et al., 2007

Similar norm effects have been documented across a variety of behaviors, including binge drinking in college students (Wechsler et al. 2003), reuse of towels in hotels (Goldstein et al. 2008), stealing petrified wood souvenirs from national parks (Cialdini et al. 2006), and staying quiet in libraries (Aarts and Dijksterhuis 2003). And this understanding has proven to be profitable as well: OPOWER, an energy start-up that focuses on presenting consumption patterns to consumers in a socially aware way, has seen its profits octuple in the past year.

2.3 Mobile persuasion

Parallel to the growth of the body of literature supporting the power of norms to influence behavior has been the rise of the mobile phone as a tool of persuasion (Fogg and Eckles 2007).

The figures on the global growth of mobile phone usage are unambiguous: mobile phones have conquered the world. Measured by volume of mobile phone subscriptions, mobile phone usage has experienced double-digit growth every year for the past decade (see Figure 2-5 on the next page). This growth rate far outstrips that of internet usage or personal computer ownership International Telecommunications Union (2010). In developed countires, we see a subscription rate of 116 subscriptions per 100 inhabitants. In the developing world, the rate is 68 subscriptions per 100 inhabitants – in other words, over two-thirds of the population of the developing world has a cell phone plan.

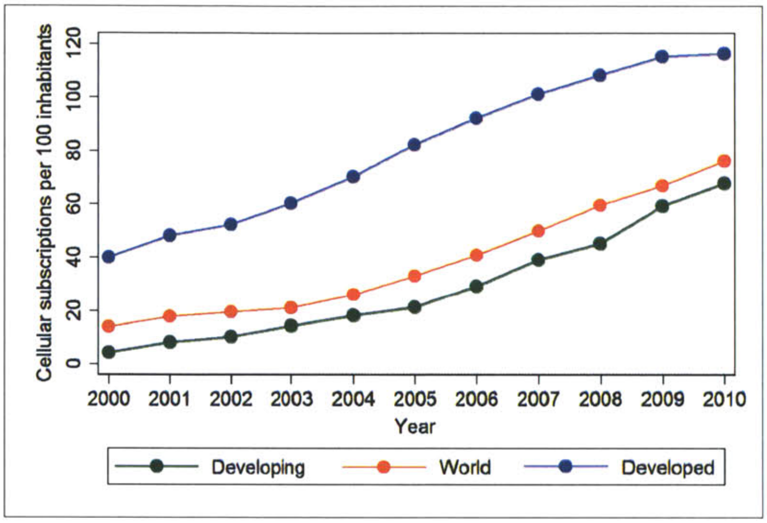

Figure 2-5: Growth in mobile phone subscriptions since 2000.
Source: ITU World Telecommunication/ICT indicators database

As for the United States, a 2010 Pew Research Center report identifies the cell phone as Americans' "gadget of choice" (Smith 2010). According to the report, 85% of adults and 75% of teenagers own at least one mobile phone. What's more, a whopping 96% of adults aged 18-29 own a mobile phone, indicating that the ownership rates for adults and teenagers will likely increase with time.

What has enabled the mobile phone to spread as it has? It is true that it is useful and convenient, allowing us to do things that are impossible otherwise, but plenty of technologies fit that description. Some might also argue that its success compared to internet or the personal computer is in its lower infrastructure and financial demands. It goes beyond merely being useful and convenient. But the spread of the mobile phone is about something deeper than that. BJ Fogg, head of the Persuasive Technology Lab at Stanford University, captures the essence of the revolution:

> Has there ever been a technology more personal and more loved than the mobile phone?... Just think about your own life: If you go out for the evening and forget your phone at home, you will probably feel anxious. You won't feel completely whole, entirely yourself. Like the love of your life, the mobile phone completes you ... We usually spend more time with

> our mobile phones than with our spouses or partners. Nothing else in the world – not even our favorite t-shirt – spends more time in our presence. (Fogg and Eckles 2007)

Perhaps Fogg is describes the relationship more romantically than the average person, but he captures the emotional element that makes the relationship with the mobile phone special. People love their phones and not only cannot imagine but really do not want to return to life without them. Most of us eagerly make our mobile phones an integral part of our lives, and optimally in the smartest form we can afford.

2.3.1 Mobile phones and behavior change

The existence of a smart, ubiquitous, portable technology that people do not see as burdensome or intrusive creates possibilities in countless fields of industry. Because it is a device that people check regularly, if not compulsively, many of these applications have focused on the potential of mobile phones to influence behavior. Health professionals have tested its usefulness as a way to remind people to take medication (World Health Organization 2003). In the developing world, mobile phones provide a better way to reach populations en masse than any other technology, creating great opportunities for improvement in provision of health care (Kaplan 2006; Fjeldsoe et al. 2009) and banking services (Mallat et al. 2004; Laforet and Li 2005; Hughes and Lonie 2007). Major technology players in private industry have launched research efforts to capitalize on applications and services on mobile phones (Kass 2007; Hedtke 2007; Consolvo et al. 2008). As we will see, the question of influencing behavior is relevant not only to businessmen and health professionals, but also to policymakers in all fields.

2.4 The Human Dynamics Group

As policymakers and researchers have turned to social norms intervention as a way to achieve compliance on the cheap, the potential role of mobile technology has not gone unnoticed. However, there are still many design questions up in the air, as well as questions about long-term effectiveness (Kaplan 2006). For the past decade, the Human Dynamics Group at the Massachusetts Institute of Technology Media Lab has combined the study of mobile technologies and the study of social influence. After running a series of experiments using mobile phones to study face to face social networks, extending the developed techniques to generating social influence seemed a natural next step.

2.4.1 The Friends and Family Study

The Friends and Family Study is the third in the Reality Mining series of mobile phone experiments run by the Human Dynamics Group, and it is the first to use couples and families as the participant pool. Partiicpants in the study are given Android smartphones with study software installed. The software captures certain phone activity as well as running periodic wifi, GPS, and bluetooth scans to establish location and proximity information. In addition to information collected by the phone, participants are required to complete daily, weekly, and monthly surveys, the content of which varied according to research demands at the time the survey is issued.

In the pilot phase of the study, which ran from March through June 2010, mobile phones with study software were handed out to 50 subjects (25 couples). In September 2010, the study expanded to 123 subjects (61 couples, along with one participant from the pilot phase whose spouse was unable to participate during the fall semester).

2.4.2 The FunFit intervention

The FunFit intervention is an experiment that was conducted on a subset of the Friends and Family subject pool. Specifically, we chose to focus on the area of health behaviors, which is a popular target for behavior interventions (Wantland et al. 2004; Tufano and Karras 2005; Abraham and Michie 2008). The design of the intervention is described in detail in the following chapter.

2.5 Summary

In this chapter, we reviewed the development of social norms literature and the rise of the study of norms as a tool of behavior change. Legal scholars have begun to take note of the power of social norms in encouraging compliance with policies. Simultaneously, people have looked to mobile phones, not so much in compliance with policies but to encourage desired behavior. The FunFit intervention of the Friends and Family Study in the Human Dynamics Laboratory seeks to combine these two threads of work to see how the power of social norms and the power of mobile phones can be combined to encourage compliance and increased levels of target behavior.

Chapter 3

FunFit Intervention Design

During the fall semester of 2010, an active intervention was carried out in the Friends and Family subject pool to explore the question of social influence using mobile phones, and further to investigate the use of the mobile phone as a tool of persuasion itself. The intervention centered on the health behavior of increasing activity levels in the study population. It was advertised to subjects as a game to help them increase their average daily activity level, and was called "FunFit." Subjects were given the option of participating, and 108 out of 123 subjects elected to participate.

3.1 FunFit timeline

FunFit ran over 62 days in the fall semester of 2010, from October 5 to December 5. The period from October 5 to October 27 was used to collect baseline activity level data from subjects. Subjects were notified of the official commencement of FunFit on October 28. From October 28 to December 5, subjects were given feedback on their performance at regular (three-day) intervals. Feedback was presented through graphics to be described in the sections that follow.

3.2 Calculation of activity level

3.2.1 Validity of accelerometer data for measuring energy expenditure

In order to study activity levels, software that captured accelerometry data was added to the phones. Throughout the past two decades, the accelerometer has been established and refined as a tool for tracking physical activity (Janz et al. 1995; Bouten et al. 1997; Eston et al. 1998; Hendelman et al. 2000; Mathie et al. 2004; Troiano et al. 2008). Although there is some error associated with using accelerometers to track energy expenditure in free-living situations, especially associated with their shortcomings in measuring sedentary activity and static exercise such as weight training (Bouten et al. 1997; Hendelman et al. 2000), a significant relationship between accelerometer output and energy expenditure has nevertheless been established (Bouten et al. 1997). In fact, accelerometer data has been found to provide more accurate estimates than other widely-used proxies for energy expenditure (Eston et al. 1998), and it has become accepted practice to use accelerometer data as a reliable proxy for energy expenditure in academic studies (Troiano et al. 2008).

3.2.2 Components of algorithm for calculation of activity level

Because one of our goals was to test the usefulness of mobile phones as a tool for investigating human behavior, we made the philosophical decision to minimize dependence on any data that could not be automatically collected by the software on the mobile phone. The algorithm for calculating activity level in the population relied solely on accelerometer data. Although versions of the algorithm which incorporated wifi location hits (to give activity points for time spent at the gym, for example) or self-report data were considered, we decided that including factors such as these would reduce portability of this algorithm to other experiments or applications. This decision enabled us to use only data which could be detected and processed on the phone. Future smaller-scale iterations of this experiment may incorporate other versions of the algorithm in order to compare accuracy.

3.2.3 Algorithm for calculation of activity level

The algorithm for calculating each person's activity score was based on accelerometer score. Accelerometer scans were taken for 15 seconds every two minutes. The score was calculated by giving one point for every second when the variance of the magnitude of acceleration was above a certain threshold. Therefore, the maximum score per reading was 15. Reward was then calculated by summing a person's scores and comparing to their average activity over a moving window of the three previous days. Reward was based on the percentage improvement over that average activity. Reward was distributed in intervals of fifty cents, from fifty cents to five dollars.

3.3 Experimental conditions

The subjects were broken up into three conditions, one control and two experimental, in order to test the effect of social influence on increasing activity level. All subjects were given feedback about their performance, although the feedback was not given directly in terms of their calculated activity level. Instead, activity level numbers were translated to a monetary reward, and subjects were shown a version of the time-series bar chart shown in Figure 3-1. Subjects were told that this graphic charted their "performance" and that reward was calculated from measured activity levels. Each subject had a unique URL which was kept up-to-date with their latest reward graphic. The URL could be accessed from a button on their phone or from any web browser. Subjects were notified by email every time their reward graphic was updated. There were no restrictions in terms of communication throughout the experiment, i.e. each subject could talk to other study subjects as he pleased.

3.3.1 Control Condition

In the control condition, subjects saw only their own progress, and their feedback was exactly as pictured in Figure 3-1. The reward given to the subjects was dependent only on their own activity.

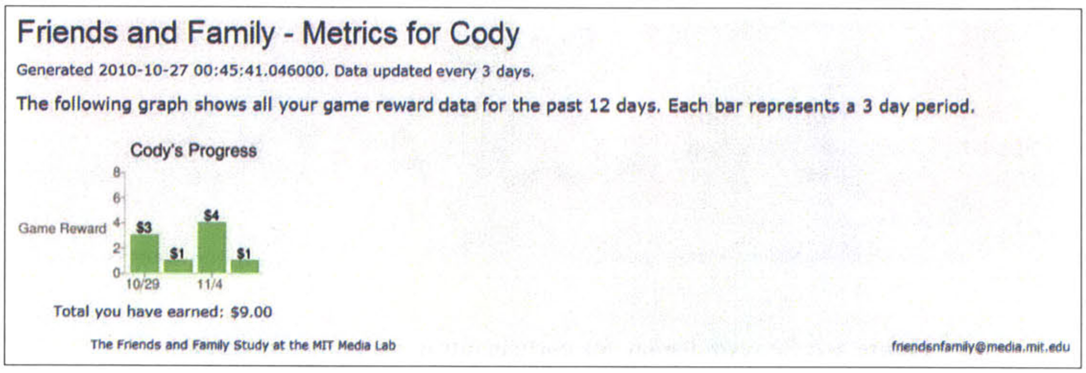

Figure 3-1: Reward display for participants in control condition

3.3.2 Experimental Condition 1

In the first experimental condition ("Experimental 1"), subjects were shown their own progress, as well as the progress of two "Buddies." Buddies were randomly assigned to each subject by the researchers. Graphics for Condition 1 subjects appeared as in Figure 3-2 on the next page. Although each subject in Experimental 1 could see the progress charts of two other subjects, his reward depended only on his own activity. Symmetrically, each subject's progress chart was also shown to two other subjects in Experimental 1. Spouses were not allowed to be paired as Buddies, but otherwise there were no restrictions on Buddy assignments.

This condition was designed following the findings of numerous studies demonstrating a significant effect on behavior from perceived group and social norms (Cial-

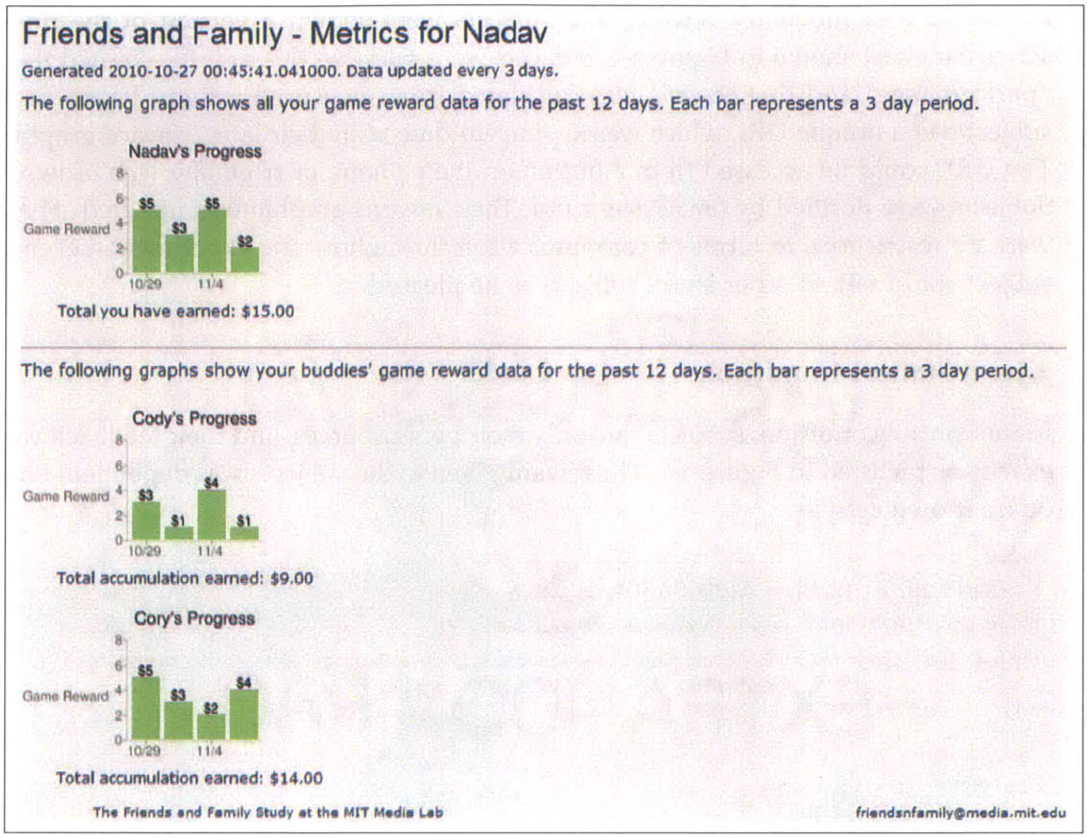

Figure 3-2: Reward display for participants in experimental condition 1

dini et al. 1990; Terry and Hogg 1996; Wechsler et al. 2003; Schultz et al. 2007). This led to our first hypothesis:

Hypothesis 1. *Activity levels of subjects in Experimental 1 will increase more than those of subjects in the control condition.*

Hypothesis 1 follows from two observations supported by findings in these social norms studies. First, people try to match the perceived normal level of an activity. This would serve to bring up the activity level of people who have higher-performing Buddies. However, studies also showed there is a potential "boomerang effect" for people performing above the average level. That is, people who are earning more reward than their Buddies (and therefore know themselves to be performing better than their Buddies) might decrease their activity level because of this knowledge. The boomerang effect can be avoided if an injunctive norm is activated in order to encourage the desired behavior. In this case, earning the financial reward served as the injunctive norm to encourage above-average performers to continue to strive for above-average results.

3.3.3 Experimental Condition 2

In the second experimental condition ("Experimental 2"), subjects were shown their own progress as well as the progress of two Buddies, as in Experimental 1. The reward mechanism, however, differed. In Experimental 2, subjects' rewards actually depended on the performance of their Buddies. The graphic displayed to subjects in Experimental 2 explained the reward mechanisms, as shown in Figure 3-3. Additionally, each subject knew that two other subjects depended on him to earn rewards for them. As in Experimental 1, spouses were not allowed to be paired as Buddies, with no other restrictions on Buddy assignments.

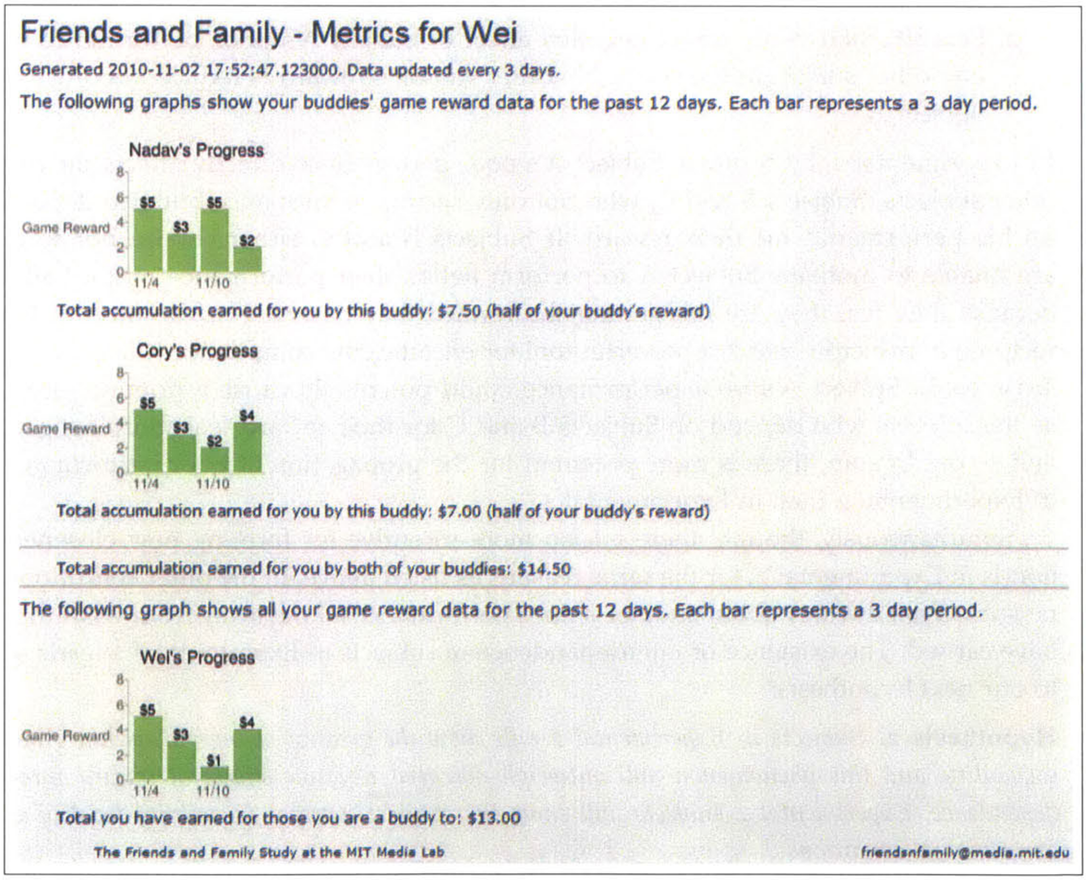

Figure 3-3: Reward display for participants in experimental condition 2

The goal of making rewards dependent on other people was to force formation of social ties. Subjects were not given explicit instructions to reach out to their Buddies, so predicting the performance of subjects in this condition was not as straightforward as for Experimental 1. First, we had to consider in Experimental 2 are the possible second-order effects of a low activity level, which do not exist in either the control condition or Experimental 1. For example, say that one person, Subject A, performs poorly. In Experimental 1, we would expect two factors to mitigate the effect of Subject A's poor performance on the group:

1. First, according to the theory in which individuals meet the average performance of the group (Schultz et al. 2007), the two other subjects who see Subject A's poor performance would only potentially be negatively affected if their performance is poorer still. However, even this potentiality is suppressed by the financial incentive for good performance, serving in the role of an injunctive norm to prevent a boomerang effect.

2. Second, there is no direct negative effect of Subject A's poor performance on any other subject in the study. No other subject's reward is reduced because of Subject A.

In Experimental 2, by contrast, Subject A's poor performance directly affects the two other subjects, Subjects B and C, who not only see his performance but also depend on his performance for their reward. If Subjects B and C either choose not to or are unable to motivate Subject A to perform better, their performance might suffer because they feel they are not getting the reward they deserve – a variation of the reciprocity principle that is a powerful tool for encouraging compliance (Cialdini and Trost 1998). Subject A's poor performance could potentially cause a domino effect, as the subjects who depend on Subjects B and C for their reward feel shortchanged, and so on. In sum, there is more potential for the propagation of poor performance in Experimental 2 than in Experimental 1.

Simultaneously, though, there is also more incentive for forming new closeness bonds in Experimental 2, for the same reasons as listed above. In the other conditions, no subject need worry about another's poor performance taking away from what they have earned. The existence of interdependence of subjects in Experimental 2 leads us to our next hypothesis:

Hypothesis 2. *Subjects in Experimental 2 will show the greatest change in within-group sociability, and this phenomenon will outweigh potential negative effects of reward interdependence. Experimental 2 subjects will show the greatest increase in activity level of all experimental conditions.*

3.4 Evaluating changes in social bonds

A further aspect of social influence could be studied by our intervention design because of preexisting relationships between the subjects. One factor which we hypothesized would affect strength of social influence is the level of closeness between two subjects. Therefore, it was important to be able to evaluate subjects' closeness. In order to study the difference in data collected by traditional social science tools to data collected by mobile phones, we devised two methods of measuring closeness.

3.4.1 Survey-based method of measuring closeness

The first method of measuring closeness relied on a traditional survey. Before the subjects were told about the FunFit intervention, we asked each subject to rate their closeness to everyone else in the study, using the scale shown in Table 3.1. Subjects

Rating	Relationship
0	I don't know this person.
1	I know of this person.
2	This person is an acquaintance.
3	This person is a friend (low ranking).
4	This person is a friend (medium ranking).
5	This person is a friend (high ranking).
6	This person is a close friend.
7	This person is family or as close to me as a family member.

Table 3.1: Relationships associated with closeness ratings.

were required to complete the same closeness ratings at the conclusion of FunFit.

3.4.2 Phone-based method of measuring closeness

To compare automatically-collected to self-reported data, we used the Bluetooth capabilities of the mobile phones carried by the study subjects. We hypothesized that Bluetooth proximity hits were a reasonable proxy for the amount of time two people spent together.

3.4.3 Effect of closeness on activity level

Because of these preexisting relationships and tools for studying closeness, another variable we will be able to investigate was the effect of closeness to one's Buddies on increase in activity level, leading to the following hypothesis:

Hypothesis 3. *Buddy triads that see the greatest increase in closeness through the course of the intervention will also see the greatest increase in activity level over the course of the intervention, closely followed by the Buddy triads who were closest at the beginning of the study.*

These closeness ratings are also relevant to Hypothesis 2 on page 34. In addition to the bluetooth proximity data discussed above, the before and after closeness ratings are a second mechanisms by which we can gauge the change in closeness of Buddy triads through the trial to test Hypothesis 2.

3.5 Summary of hypotheses

Taken together, these hypotheses will help us answer the central thesis questions proposed in Section 1.4. Hypothesis 1 on page 32 allows us to answer whether mobile phones as a tool for studying social influence questions can duplicate established results. Hypothesis 2 on page 34 allows us to look at the question of how to combine mobile phone-based persuasion and social influence-based persuasion effectively. Hypothesis 3 on the preceding page allows us to look at both questions, first by allowing us to compare survey- and phone-collected data, and second by again exploring the effect of the combined persuasion.

Chapter 4

Results

Data collected by the mobile phones was uploaded to a secure server housed at the MIT Media Lab. The closeness survey to establish pre-existing relationships was administered online, and responses were downloaded and stored in the same database as the data collected by the phones. This analysis focuses on the accelerometer readings and pre-intervention closeness scores.

Eleven subjects were removed from the study pool over the course of the intervention. Their data has been totally removed from the analysis. For details on the number of subjects in each study condition, see Table 4.1. More people were placed in the experimental conditions because of the triadic Buddy groups.

Condition	Initial	Dropped	Total
Control	18	2	16
Experimental 1	45	5	40
Experimental 2	45	4	41

Table 4.1: Number of subjects in each condition

4.1 Analysis of activity levels

Using the accelerometer data gathered by the phones, we first determined the changes in activity level throughout the study period of October 5-December 5. (Refer to Section 3.1 on page 29 for a description of the timeline of the intervention.)

4.1.1 Daily average activity level

Daily average activity levels were calculated by summing all accelerometer scores 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 scans were taken every two minutes for a total of 720 possible readings per day per individual. Multiplying the daily average activity level by 720 would give the total activity score for the day but would not result in qualitatively different calculations. Daily average activity levels can be calculated for different units of analysis. For example, on a given day, a daily average activity score could be calculated for an individual by summing all his scores for the day and dividing by his total count of accelerometer readings for the day. On teh same day, a daily average activity score could also be calculated for the control group by summing all the scores of all subjects in the control group and dividing by the total count of accelerometer readings gathered from all subjects in the control group.

Figure 4-1 shows the distribution of daily average activity levels, calculated by condition (Control, Experimental 1, and Experimental 2). That is, for one "daily average activity level," all the day's readings for every subject in a given condition are summed and normalized, resulting in 3 conditions times 62 days = 186 daily average activity level readings. The distribution appears normal and is confirmed to be normal by a chi-squared test (a p-value of 0.6192 when testing the null hypothesis of normality, indicating that the hypothesis should not be rejected).

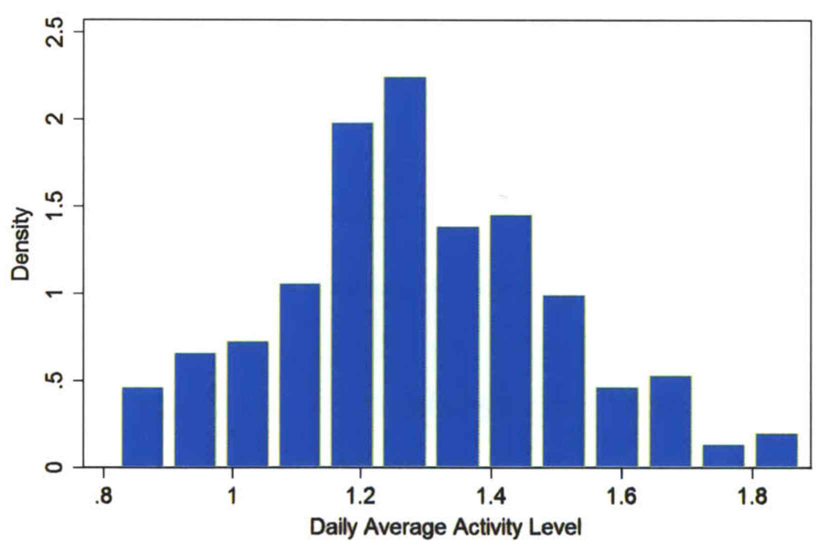

Figure 4-1: Histogram of daily average activity levels

4.1.2 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 4.2.

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

Table 4.2: For the analysis of changes in activity level, the intervention is broken up according to the periods shown above.

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. In other words, the pattern of the increase in activity is important. A group of subjects that shows a steady increase through Periods 2 and 3 may be qualitatively different than a group of subjects that shows an activity level spike in Period 2 but drops off again in Period 3.

4.1.3 Activity levels by condition

Tables 4.3 on the next page and 4.4 on page 41 present information about daily average activity levels. Table 4.3 compares the average activity levels between groups and within periods, and Table 4.4 compares the activity level within groups and across periods. In both tables, additional data is presented regarding activity levels with the experimental groups divided according to the pre-intervention closeness of the Buddy triads, as declared on the closeness survey. The number of subjects in each experimental group when divided by closeness can be found in Table 4.5 on page 41.

The Kolmogorov-Smirnov (K-S) test is applied for all significance testing of activity level differences. The K-S test was chosen because of its insensitivity to the shape of the distributions being tested. This was necessary because, despite the normality of the distribution of daily average activity levels when all subjects are combined, the normality test is not passed in all cases when the calculated activity levels are divided by experimental condition. In the K-S test, the null hypothesis being tested is that the two distributions being compared are equal. A low p-value indicates that the null hypothesis should be rejected, i.e. the distributions are not equal.

Groups being tested	Group 1 mean	Group 2 mean	D	p-value
Pre-Intervention (Period 1)				
Control vs. Exp 1 & 2	1.200	1.290	0.3261	0.046*
Control vs. Exp 1	1.200	1.295	0.3478	0.078
Control vs. Exp 2	1.200	1.284	0.3478	0.078
Exp 1 vs. Exp 2	1.295	1.284	0.1739	0.816
Close Buddies:				
Exp 1 vs. Exp 2	1.568	0.988	0.4048	0.558
Stranger Buddies:				
Exp 1 vs. Exp 2	1.372	1.266	0.3333	0.630
Mixed Buddies:				
Exp 1 vs. Exp 2	1.050	1.429	0.3750	0.516
Post-Intervention (Periods 2 and 3)				
Control vs. Exp 1 & 2	1.215	1.337	0.3718	0.001***
Control vs. Exp 1	1.215	1.355	0.4103	0.001***
Control vs. Exp 2	1.215	1.320	0.3333	0.015*
Exp 1 vs. Exp 2	1.355	1.320	0.1282	0.862
Close Buddies:				
Exp 1 vs. Exp 2	1.590	1.340	0.7778	0.036*
Stranger Buddies:				
Exp 1 vs. Exp 2	1.328	1.315	0.1190	0.998
Mixed Buddies:				
Exp 1 vs. Exp 2	1.327	1.313	0.2143	0.878

* $p < 0.05$

*** $p < 0.005$

Table 4.3: The significance in the differences between the means of the various conditions are compared for the pre- and post-intervention periods. 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		Experimental 1		Experimental 2	
Period	Activity	Δ	Activity	Δ	Activity	Δ
Overall						
1	1.200	—	1.295	—	1.284	—
2	1.225	+0.025	1.363	+0.068	1.296	+0.012
3	1.205	+0.005	1.346	+0.051*	1.342	+0.058
Close Buddies (both Buddies score 3 or higher)						
1			1.568	—	0.988	—
2			1.639	+0.071	1.228	+0.240
3			1.685	+0.117	1.235	+0.247**
Stranger Buddies (both Buddies score 2 or lower)						
1			1.372	—	1.266	—
2			1.299	-0.073	1.355	+0.089
3			1.354	-0.018	1.363	+0.097
Mixed Buddies (one Close, one Stranger)						
1			1.050	—	1.429	—
2			1.351	+0.301	1.290	-0.139
3			1.182	+0.132*	1.383	-0.046

* $p < 0.05$
** $p < 0.01$

Table 4.4: Mean activity level by condition and period, overall and divided by closeness of Buddy triads. (The control condition, in which there were no Buddies, does not vary throughout the three closeness levels.) Note that the differences in activity level Δ are calculated from Period 1, so that the difference in Period 3 is the activity level in Period 3 minus the activity level in Period 1, not minus the activity level in Period 2. Significant differences are indicated; p-values are taken from a K-S test. Only differences between Period 3 and Period 1 (the bottom line in each panel) were tested for significance.

	Experimental 1	Experimental 2
Close Buddies	6	9
Stranger Buddies	22	12
Mixed Buddies	12	20

Table 4.5: Number of subjects in each condition when divided by closeness categories

Although subjects were assigned randomly to their conditions, Table 4.3 on page 40 shows a significant difference in the pre-intervention activity levels of the control versus experimental groups (the two experimental groups are combined in this instance). The two experimental groups are not significantly different from each other in any of the closeness conditions. In the post-intervention period, the control group is again significantly less active than the combined experimental groups, but now it is also significantly less active than either of the two experimental groups on its own. The two experimental groups differ significantly only when comparing Buddy triads with close bonds before the intervention began. In this case, Experimental 1 is significantly more active than Experimental 2.

Table 4.4 on the previous page shows daily average activity level subdivided by three categories: experimental condition, time period, and pre-existing closeness to Buddies. Additionally, the change in activity level from Period 1 is presented for Periods 2 and 3. Activity level differences between Period 1 and Period 3 were tested for significance. Period 3, rather than a combined score from Periods 2 and 3, was chosen for the comparison on the reasoning that the activity level at the end of the intervention was most representative of the effect of the intervention. The activity level increase of subjects in Experimental 1, the condition in which subjects see others' data but earn their own reward, show a significant difference overall in activity level. We also see a significant increase for Experimental 2 in the Close Buddies condition and a significant increase for Experimental 1 in the Mixed Buddies condition.

Overall, the experimental conditions with the Buddies show a greater increase in activity level through the course of the intervention than the control group. It is interesting to note that the difference in how activity level increases in the two experimental conditions. In Experimental 2, where each subjects depends on others for his reward, the activity level increases steadily throughout the course of the intervention for three of the four Buddy conditions. The exception is the Mixed Buddies condition. By contrast, in Experimental 1, where each subject can see the data of two other people but gets rewarded based on his own activity, activity between the three periods is less predictable.

4.1.4 Reward efficiency

To measure the effectiveness of our incentives, we look at what we call "reward efficiency." Reward efficiency measures the cost of each unit increase in activity. As discussed in Section 3.2.3 on page 30, the reward that was given to the subjects for increasing activity level followed the general trend of activity. However, because the reward came in discrete chunks of fifty cents, it could not track activity perfectly. This resulted in some variation in the amount of activity that could be performed for a certain reward, which allowed us to look at the efficiency of the reward given for each group. (If reward had tracked activity perfectly, there should be no difference in cost per dollar in the different groups.)

Reward efficiencies are calculated using average reward amount in Period 3, divided by the difference in activity level between Period 3 and Period 1 (see Table 4.6).

Condition	Activity Change from Period 1 to Period 3	Reward in Period 3	Reward Efficiency (Δ/\$)
Overall			
Control	0.005	\$2.99	0.0016
Exp 1	0.051	2.76	0.0185
Exp 2	0.058	3.00	0.0193
Close Buddies (both Buddies score 3 or higher)			
Exp 1	0.117	\$2.64	0.0443
Exp 2	0.247	2.96	0.0834
Stranger Buddies (both Buddies score 2 or lower)			
Exp 1	-0.018	\$2.80	-0.0064
Exp 2	0.097	2.88	0.0337
Mixed Buddies (one Close, one Stranger)			
Exp 1	0.132	\$2.64	0.0500
Exp 2	-0.046	3.10	-0.0148

Table 4.6: Reward efficiency is defined as the amount of activity level increase per dollar of reward paid.

K-S tests were also performed to do a pairwise comparison of the differences in reward efficiency between groups (Table 4.7 on the following page. All of the differences tested are significant, except for the reward efficiencies of Experimental 1 versus Experimental 2 when not divided by Buddy conditions.

As with the activity levels, the reward efficiency of the control condition is low compared to the experimental conditions. Here, the reward efficiency is significantly lower for the control condition than either of the experimental conditions. For every dollar we pay a subject with the incentive structure of the control condition, we get less activity than with a subject with the incentive structure of either of the experimental conditions. The experimental conditions are approximately equal in the overall analysis but show more variation when divided into Buddy categories.

4.2 Days with missing data

In performing the activity level analysis, we noticed that days with missing data had some features which might affect our analysis. For one person, a complete day's

Groups being compared	Group 1 reward efficiency	Group 2 reward efficiency	D	p-value
Overall				
Control vs. Exp 1	0.0016	0.0185	1.0000	0.001**
Control vs. Exp 2	0.0016	0.0193	1.0000	0.001**
Exp 1 vs. Exp 2	0.0185	0.0193	0.5476	0.200
Close Buddies (both Buddies score 3 or higher)				
Exp 1 vs. Exp 2	0.0443	0.0834	1.0000	0.002**
Stranger Buddies (both Buddies score 2 or lower)				
Exp 1 vs. Exp 2	-0.0064	0.0337	01.0000	0.002**
Mixed Buddies (one Close, one Stranger)				
Exp 1 vs. Exp 2	0.0500	-0.148	1.0000	0.008**

* $p < 0.05$
** $p < 0.01$

Table 4.7: The significance in the differences between the reward efficiencies of different conditions in Period 3. All differences are significant, except for the difference between the two experimental groups when taken in their entirety.

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 we did not have any reading for that interval. We can think of missing data as representing the phone being off.

Condition	Missing
Control	29.8%
Condition 1	26.3%
Condition 2	27.9%

Table 4.8: Amount of missing data for each condition. Missing data is calculated as a percentage of missing readings out of total possible readings (720 readings per day x 62 days x number of subjects in the condition)

Missing data is approximately evenly distributed between groups (see Table 4.8), but the distribution of missing data among both days and time of day is uneven. Because our analysis is built around data aggregated by day, it was important to explore if this uneven distribution might affect our daily average activity level calculations. The algorithm for calculating the daily average activity level, as discussed in

Section 4.1.1 on page 37, was to sum all the day's scores for our unit of analysis (in most cases, one of the study groups) and divide this figure by the total number of the day's readings, added over the same unit of analysis. This method assumes that any readings missing from that day's sum are not systematically biased.

We first established that the average accelerometer score per reading would vary by time of day. This made sense because we would expect more zero and low readings while people are sleeping, for example, than in the middle of the afternoon. A visual inspection of the distribution of non-zero readings indicated that we should 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 4.9 confirms that activity varies greatly throughout the day.

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 4.9: 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.

This uneven distribution of activity over the day could affect our calculation of daily activity level for days with a high proportion of missing data when combined with two other features:

1. First, that on days where an individual is missing a high percentage of data, the non-missing data is not evenly distributed throughout the day (for example, imagine we get all the readings from the noon-6PM period in Table 4.9; then the average activity level for that day as we calculate it will be higher than if we had all the data).

2. Second, that on days where an individual is missing a high percentage of data, the other subjects in his condition cannot make up for that individual's skew on that day – that is, the missing people-days themselves are not randomly distributed throughout the experiment but rather cluster on certain days. Then, rather than one person's skew being mitigated by others, it is instead exaggerated.

Our dataset exhibits both of these features. Table 4.10 on the next page shows that, on days with fewer than 180 readings (i.e., less than 25% of the maximum 720 readings), 62.5% of the readings come from the 6PM-midnight period. In contrast, on days with more than 540 readings (i.e., more than 75% of the maximum 720 readings), the data are evenly distributed thoughout the day, ad 24.1% of readings come from the 6PM-midnight period. The implication from these numbers is that, if we were to calculate daily activity on people-days with less than 25% of data collected using the same method we used in our analysis, the activity level calculated would be higher than on days where there is a low fraction of missing data.

Time of Day	Number of Readings	% of Total Readings
Days with < 25% Data Collected		
Midnight-6AM	2990	15.1
6AM-Noon	1554	7.8
Noon-6PM	2908	14.6
6PM-Midnight	12404	62.5
Days with 25-50% Data Collected		
Midnight-6AM	26724	18.4
6AM-Noon	25201	17.4
Noon-6PM	42048	28.9
6PM-Midnight	51232	35.3
Days with 50-75% Data Collected		
Midnight-6AM	110843	21.9
6AM-Noon	120226	23.7
Noon-6PM	139763	27.6
6PM-Midnight	135581	26.8
Days with > 75% Data Collected		
Midnight-6AM	622270	25.8
6AM-Noon	608444	25.2
Noon-6PM	602310	24.9
6PM-Midnight	581855	24.1

Table 4.10: Distribution of readings among quarters of the day, for all days with less than 25% of data collected.

Similarly, our data has the second feature of non-randomly-distributed missing days. Missing people-days tended to cluster in the pre-intervention period (Figure 4-2 on the facing page). We would thus expect our daily average activity level calculations for the pre-intervention period to be higher than if there were no missing data.

It is important to note that these days are characterized by different subject behavior, but it is not known how this reflects on actual activity for the day. When subjects do not use their phone much, they tend to use it in the evening, but nothing can be said about their behavior for the rest of the day. However, it is plausible

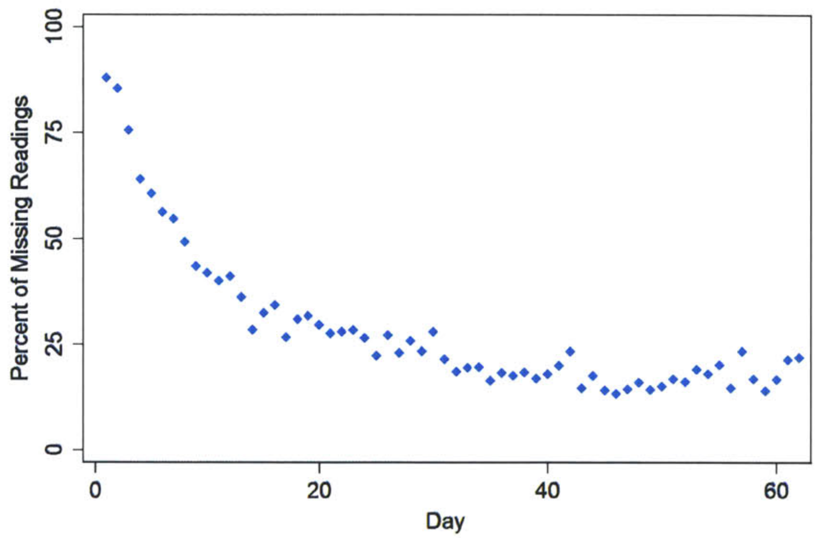

Figure 4-2: Percent of missing data versus day

that our method for translating the data into an activity level would be sensitive to this tendency. This might not affect our between-group analysis, as missing data is distributed evenly between groups (see 4.8 on page 44), but it could affect our between-period analysis.

4.3 Activity analysis repeated with missing data removed

We decided to drop all people-days with fewer than 50% of the possible 720 readings and rerun the analysis. That is, if any subject had fewer than 360 readings on a given day, all the readings for that subject for that day were not used in any analysis. We decided on this threshold because it was after the 50% point that the skew in data readings by time of day seemed to even out, as no quarter day accounted for more than 28% of the data (Table 4.10 on the facing page. The following section will go through the same steps as were performed in 4.1 on page 37.

4.3.1 Daily average activity level

Figure 4-3 on the following page shows the distribution of daily average activity levels with days missing > 50% of readings removed. The distribution appears normal and is confirmed to be normal by a chi-squared test (a significance level of 0.9270 when testing the null hypothesis of normality).

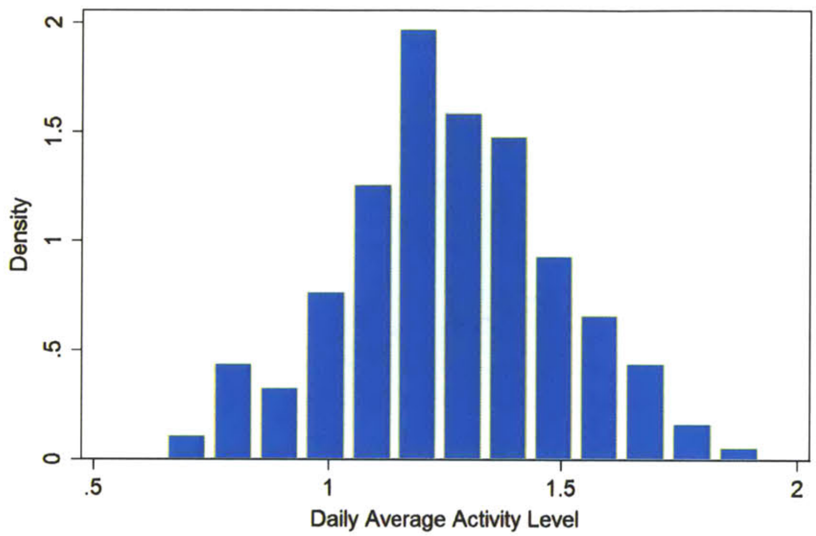

Figure 4-3: Histogram of daily average activity levels, days with > 50% missing data removed

4.3.2 Activity levels by condition

Tables 4.11 on the next page and 4.12 on page 50 present information about daily average activity levels. Table 4.11 compares average activity level between groups and within periods, and Table 4.12 compares activity level between periods and within groups. For Experimental 1 and Experimental 2, data is also presented regarding activity level divided according to the pre-intervention closeness of the Buddy triads within the experimental condition. As in the previous analysis with the complete dataset, all significance testing is performed with the K-S test, because normality cannot be guaranteed for all subdivisions of the data.

Comparing Table 4.11 on the next page and Table 4.3 on page 40, the trends in the data look the same. The one major change is that when the data for days missing over 50% of readings are removed, the significance in the difference between the performance of Experimental 1 versus Experimental 2 in the post-intervention, close Buddies condition is lost. However, the data exhibit the trend we expected in the pre-intervention period. We expected that the activity level calculations in the pre-intervention period would skew high because of the features we identified in Section 4.2 on page 43. In fact, when the data are removed, all the pre-intervention means decrease. There is no equivalent effect on the post-intervention means.

Although our first hypothesized effect is confirmed, Table 4.12 on page 50 does not show an increased significance in the within-group, between-periods analysis.

Groups being tested	Group 1 mean	Group 2 mean	D	p-value
Pre-Intervention (Period 1)				
Control vs. Exp 1 & 2	1.162	1.241	0.3261	0.046*
Control vs. Exp 1	1.162	1.266	0.3478	0.078
Control vs. Exp 2	1.162	1.216	0.3043	0.164
Exp 1 vs. Exp 2	1.266	1.216	0.2609	0.316
Close Buddies:				
Exp 1 vs. Exp 2	1.545	0.966	0.7500	0.775
Stranger Buddies:				
Exp 1 vs. Exp 2	1.349	1.238	0.3056	0.613
Mixed Buddies:				
Exp 1 vs. Exp 2	1.021	1.321	0.3000	0.660
Post-Intervention (Periods 2 and 3)				
Control vs. Exp 1 & 2	1.207	1.328	0.3718	0.001***
Control vs. Exp 1	1.207	1.341	0.4193	0.001***
Control vs. Exp 2	1.207	1.316	0.3590	0.007**
Exp 1 vs. Exp 2	1.341	1.316	0.1026	0.976
Close Buddies:				
Exp 1 vs. Exp 2	1.313	1.229	0.5000	0.261
Stranger Buddies:				
Exp 1 vs. Exp 2	1.350	1.336	0.2863	0.517
Mixed Buddies:				
Exp 1 vs. Exp 2	1.330	1.351	0.2000	0.951

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

Table 4.11: The significance in the differences between the means of the conditions within a certain period are compared, for data with days missing > 50 % of data removed. For this analysis, the two post-intervention periods are combined into one. 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.

Period	Control		Experimental 1		Experimental 2	
	Activity	Δ	Activity	Δ	Activity	Δ
Overall						
1	1.162	—	1.266	—	1.216	—
2	1.215	+0.053	1.346	+0.080	1.289	+0.073
3	1.199	+0.037	1.336	+0.070*	1.342	+0.126
Close Buddies (both Buddies score 3 or higher)						
1			1.545	—	0.966	—
2			1.636	+0.091	1.241	+0.275
3			1.663	+0.138	1.235	+0.269**
Stranger Buddies (both Buddies score 2 or lower)						
1			1.349	—	1.238	—
2			1.289	-0.060	1.359	+0.121
3			1.342	-0.007	1.375	+0.137*
Mixed Buddies (one Close, one Stranger)						
1			1.021	—	1.321	—
2			1.313	+0.292	1.266	-0.055
3			1.175	+0.154*	1.374	0.108

* $p < 0.05$
** $p < 0.01$

Table 4.12: Mean activity level by condition and period, overall and divided by closeness of Buddy triads, for data with days missing > 50 % of data removed. The differences in activity level Δ are calculated from Period 1, so that the difference in Period 3 is the activity level in Period 3 minus the activity level in Period 1, not minus the activity level in Period 2. Significant differences are indicated; p-values are taken from a K-S test. Only differences between Period 3 and Period 1 (the bottom line in each panel) were tested for significance.

However, we do now have a significant effect in Experimental 2 for two of the Buddy conditions.

4.3.3 Reward efficiency

Tables 4.13 and 4.14 on the following page present information on reward efficiency for this dataset. The reward efficiency of the control condition improves by a factor of ten as compared with the full dataset, but it is still significantly lower than the reward efficiencies of the experimental groups.

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 4.13: Reward efficiencies for data with days missing > 50 % of data removed. Reward efficiency is defined as the amount of activity level increase per dollar of reward paid.

4.4 Summary

We use two metrics to estimate the effectiveness of our structures for eliciting behavior change. The first is to look at change in activity level directly, both across groups and across time periods, and the second is to calculate the efficiency of reward in each of the experimental groups.

When daily average activity levels are analyzed, the data provide support for a significant difference between the control group on the one hand and the two ex-

Groups being compared	Group 1 reward efficiency	Group 2 reward efficiency	D	p-value
Overall				
Control vs. Exp 1	0.0120	0.0253	1.0000	0.001**
Control vs. Exp 2	0.0120	0.0416	1.0000	0.001**
Exp 1 vs. Exp 2	0.0253	0.0416	0.4286	0.432
Close Buddies (both Buddies score 3 or higher)				
Exp 1 vs. Exp 2	0.0444	0.0896	1.0000	0.002**
Stranger Buddies (both Buddies score 2 or lower)				
Exp 1 vs. Exp 2	-0.0025	0.0464	01.0000	0.001**
Mixed Buddies (one Close, one Stranger)				
Exp 1 vs. Exp 2	0.0560	0.0171	1.0000	0.001**

* $p < 0.05$
** $p < 0.01$

Table 4.14: The significance in the differences between the reward efficiencies of different conditions in Period 3, for data with days missing > 50 % of data removed. All differences are significant, except for the difference between the two experimental groups when taken in their entirety.

perimental groups, taken together, on the other. Analysis of the difference of effect between the two 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 show more significance than the differences in activity level.

In addition to using two metrics, we also conducted analyses with two versions of the dataset: one full version, and one where people-days missing over 50% of accelerometer readings removed. The analysis confirmed our hypothesis that people-days with 50% missing data were skewing our pre-intervention figures higher, but the major trends in significance were the same as with the full version of the dataset.

Chapter 5

Discussion

5.1 General observations

Overall, all methods of analysis show a significant difference in the activity levels of the control versus experimental conditions. This difference offers strong support for a social influence effect on the two experimental groups, subject to the caveat described in Section 5.1.1. There is no evidence for a difference between the performance of the two experimental conditions when all subjects in each condition are considered. However, certain trends are suggested when experimental conditions are divided according to pre-existing closeness of the Buddy triads.

The sometimes surprising results of the signfiicance tests indicate some abnormality with the data, which has not been investigated here but which should be explored. One example can be found in Table 4.3 on page 40. In the pre-intervention period, the distributions of Experimental 1 and Experimental 2 in the Close Buddies condition are found to be equal, despite means of 1.568 and 0.988, respectively. In contrast, the post-intervention distributions of the two groups in the same Close Buddies condition are found to be significantly different, this time with means of 1.590 and 1.340. Because the distributions show some unexpected behavior, we look at trends in the data in addition to results which are found to be significant.

5.1.1 Baseline difference in activity between control and experimental groups

The major caveat regarding the significance of the difference between the results for the control group versus the two experimental groups is the existence of a significant difference in baseline activity levels for the control versus experimental groups. Even though the difference in daily average activity level becomes much more significant in the post-intervention period, it could be argued that those who are less active are less likely to become more active, even in response to the same amount of influence. In other words, even though we see a bigger difference in the post-intervention period, this could be the result of characteristics of the experimental groups rather than the result of our influence techniques. Subjects were assigned randomly to conditions, so this is an unfortunate characteristic of the random assignment. It also suggests that our control group was too small. The control group of 16 people is much smaller than either of the two experimental conditions (see Table 4.1 on page 37). This was a deliberate design choice in order to compensate for the placement of subjects in the experimental conditions into Buddy triads, which opened the possibility of further subdividing the experimental groups (as we did in the analysis). We made the experimental conditions larger in anticipation of subdividing the conditions in our analysis, but this issue should be revisited in any future iterations of the experiment.

5.1.2 Results based on closeness divisions

The random distribution of subjects to conditions also produces a quirk when experimental conditions are split according to closeness of Buddy triads. In this case, triads are unevenly distributed into each closeness condition, both within and across experimental groups (see Table 4.5 on page 41). For example, Experimental 1 has a much higher number of subjects who are in triads with two strangers than Experimental 2 does. It is not immediately obvious what the effects of this distribution might be, but it should be kept in mind when considering results based on closeness divisions.

5.1.3 Results with missing-data modification

We found that people-days with a high percentage of missing readings had different characteristics than days with a low percentage of missing readings. People-days with a high percentage of missing readings tended to have their readings cluster in the latter half of the day, rather than being evenly distributed throughout the day. Because readings from the latter half of the day had higher average accelerometer scores, and furthermore because these people-days with missing data were concentrated in the pre-intervention period, we hypothesized that these days were bringing up the daily average activity levels in the pre-intervention period. This turned out to be true, but the overall trends and significant results did not change between analyses.

5.2 Hypothesis 1

Hypothesis 1 on page 32, which states that activity levels of subjects in Experimental 1 will increase more than those of subjects in the control condition, is strongly supported by the results. In both runs of the analysis, with the two versions of the dataset, the post-intervention activity level of Experimental 1 is significantly higher than that of the control group at the $p = .001$ level (see Tables 4.3 on page 40 and 4.11 on page 49). Additionally, Experimental 1 is the only condition to show a significant increase in activity level from Period 1 to Period 3 in the overall analysis, before dividing by closeness of Buddy triads (see Tables 4.4 on page 41 and 4.3 on page 40). This provides strong evidence that mobile telephony can be used to activate existing social influence networks in support of a specific behavior change goal.

5.3 Hypothesis 2

Hypothesis 2 on page 34 contains two predictions: first, that subjects in Experimental 2 will show the greatest change in within-group sociability, and second, that subjects in Experimental 2 will show the greatest increase in activity level of all experimental conditions. The Bluetooth proximity and post-intervention closeness survey data which will be used to determine changes in sociability was not available in time for this analysis, so the first prediction cannot be assessed. The second prediction, that Experimental 2 will show the greatest increase in activity level, is not supported by the data.

Hypothesis 2 expresses our expectation that the influence structure of Experimental 2 would produce the best results. In Section 3.3.3 on page 33, we discuss how the reward structure in Experimental 2, where each subject's reward is dependent on two others, is a double-edged sword. On the one hand, this structure encourages formation of social bonds. On the other hand, it opens up the possibility of a domino effect of poor performance if existing or newly generated social bonds are not strong enough to counteract the reciprocity principle in the face of a poor earner. It is possible that the poor performance had a stronger effect than we anticipated. Again, the data to judge the formation and strengthening of social bonds was not available in time for this analysis, so the question remains open.

However, even though Hypothesis 2 was not shown to be true in terms of activity level, Experimental 2 produces the best results in other measures. The reward efficiency for the Close Buddies and Stranger Buddies conditions are significantly higher for Experimental 2 than Experimental 1 (see Tables 4.6 on page 43 and 4.13 on page 51). (The Mixed Buddies condition is discussed below.)

Trends in the data provide evidence that Experimental 2 may be the best long-run performer. In Tables 4.4 on page 41 and 4.12 on page 50, only Experimental 1 shows significance in the change in activity level between Period 1 and Period 3. However, Experimental 2 is the only group that shows steady increase through periods in the overall analysis. In fact, in Table 4.4, Experimental 2 shows a steady increase in three of the four categories (the Overall, Close Buddies, and Stranger Buddies analyses), whereas Experimental 1 shows a steady increase only in the Close Buddies analysis. In the analysis with the version of the dataset where people-days with greater than 50% missing data are removed, Experimental 2 drops off a bit in the Close Buddies analysis, but the overall effect of Experimental 2 showing a more steady upward rise across periods is still evident (see Table 4.12).

One last indication of the strength of the influence structure of Experimental 2 can be found in Table 4.12. It is true that activity level does not increase significantly from Period 1 to Period 3 overall for Experimental 2. However, once the groups are split according to Buddy closeness, Experimental 2 achieves significant increase in two of the three Buddy categories. Triads show significant increase in the Close Buddies and Stranger Buddies conditions, but not in the Mixed Buddies condition. One possible explanation for this is that the incentive to form social ties works better when both Buddies are strangers, and the social obligation incentive works better when both Buddies are close. This explanation will be tested with the Bluetooth proximity and post-intervention closeness data, which can give us insight into how social bonds changed through the course of the intervention.

5.4 Hypothesis 3

Hypothesis 3 on page 35 postulates a directly proportional relationship between changes in within-triad closeness and changes in activity level. It states that Buddy triads that see the greatest increase in closeness through the course of the intervention will also see the greatest increase in activity level over the course of the intervention, followed by Buddy triads who were closest at the beginning of the study. Although there is evidence that Close Buddy triads performed better than others in Experimental 2 (see Tables 4.4 on page 41 and Tables 4.3 on page 40, the substance of this hypothesis cannot be tested without the Bluetooth proximity and post-intervention closeness data.

5.5 Implications for behavior change

The results from the FunFit intervention show that mobiile telephony can be successfully used to elicit behavior change but works best when combined with social influence. This is a significant finding, as many of the mobile phone applications designed for behavior change do not include a social influence component (Hedtke 2007; Kass 2007; Consolvo et al. 2008). This social influence can be manufactured in the sense that it increases the salience of existing norms. Cialdini et al. (1990) establishes that salience is critical for a norm to have an effect. This is the situation in Experimental 1, where subjects are shown feedback on the performance of other subjects but rewarded based on their own activity. The feedback increases the salience of a norm to increase activity level, with significant effect.

Additionally, these results provide support for the efficacy of more explicitly manufactured influence in generating behavior change. The reward structure of Experimental 2 was designed to force formation of new social ties, and there is evidence that Experimental 2 exhibits a more steady and sustained increase in activity level. Although not all of these results show significance, the indicators are promising, and similar experiments should be repeated with increased sample sizes in order to explore this potential. This finding extends beyond mobile telephony-based interventions and can be applied to intervention techniques more generally.

Chapter 6

Conclusion

6.1 Contributions

This paper explored the effect of using mobile telephony and manufactured social influence in combination to elicit behavior change. In exploring this effect, we make the following contributions:

1. Designing a longitudinal experiment to study the effectiveness of combining mobile telephony and manufactured social influence to achieve behavior change.

2. Demonstrating that mobile telephony can be used as a tool to quantify social influence on behavior, addressing a weakness in social psychology literature.

3. Confirming the efficacy of mobile telephony in eliciting behavior change.

4. Confirming of the efficacy of social influence, both existing and generated, in eliciting behavior change.

5. Finding that the efficacy of mobile telephony in eliciting behavior change is enhanced when used in conjunction with social influence.

6. Finding that generated social ties may be more powerful than existing social ties in eliciting behavior change.

6.2 Future work

6.2.1 Changes in social bonds

One major area of this experimental setup that is not explored in this paper is the change in social bonds throughout the course of the intervention. The Bluetooth proximity and exit survey data, containing self-reported closeness to other subjects after the intervention, can be used to trace the evolution of closeness in the community through the course of the intervention. This data may help us understand the distribution abnormalities mentioned in Section 5.1 on page 55. Analysis of this data will allow us to complete evaluation of Hypotheses 2 and 3 proposed in Chapter 3.

6.2.2 Persistence of changes

A key point of interest when conducting interventions is the persistence of behavior change after the intervention period. Interventions can have a high cost, in terms of time, money, and effort, and it is impossible to carry on interventions indefinitely. Therefore, persistence of behavior change elicited by any intervention is critical. The intervention period was divided into two smaller periods to take a first look at persistence, and there is reason to believe the influence structure in Experimental 2 generated more lasting results. Though the FunFit intervention has officially ended, the phones used in the study are still being carried by the subjects, and so longer-term persistence can also be explored.

Bibliography

Aarts, Henk, and Ap Dijksterhuis. 2003. The silence of the library: Environment, situational norm, and social behavior. *Journal of Personality and Social Psychology* 84(1):18–28.

Abraham, Charles, and Susan Michie. 2008. A taxonomy of behavior change techniques used in interventions. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association* 27(3):379–87.

Ajzen, Icek. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50(2):179–211.

———. 2002. Perceived behavioral control, self-ffficacy, locus of control, and the Theory of Planned Behavior. *Journal of Applied Social Psychology* 32(4):665–683.

Ajzen, Icek, and Martin Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Armitage, Christopher J., and Mark Conner. 2001. Efficacy of the Theory of Planned Behaviour: a meta-analytic review. *The British journal of social psychology / the British Psychological Society* 40(Pt 4):471–99.

Bandura, Albert. 1982. Self-efficacy mechanism in human agency. *American Psychologist* 37(2):122–147.

———. 1989. Human agency in social cognitive theory. *American Psychologist* 44(9): 1175–84.

———. 1997. *Self-efficacy: the exercise of control*. W. H. Freeman and Company.

Bohner, Gerd, and Nina Dickel. 2011. Attitudes and attitude change. *Annual review of psychology* 62:391–417.

Bouten, Carlijn V. C., Karel T. M. Koekkoek, Maarten Verduin, Rens Kodde, and Jan D. Janssen. 1997. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE transactions on bio-medical engineering* 44(3):136–47.

Cialdini, Robert B., Linda J. Demaine, Brad J. Sagarin, Daniel W. Barrett, Kelton Rhoads, and Patricia L. Winter. 2006. Managing social norms for persuasive impact. *Social Influence* 1(1):3–15.

Cialdini, Robert B., and Noah J. Goldstein. 2004. Social influence: compliance and conformity. *Annual Review of Psychology* 55:591–621.

Cialdini, Robert B., Raymond R. Reno, and Carl A. Kallgren. 1990. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology* 58(6):1015–1026.

Cialdini, Robert B., and Melanie R. Trost. 1998. Social influence: Social norms, conformity, and compliance. In *The handbook of social psychology, volume 2*, ed. Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey, 151–192. McGraw-Hill.

Conn, Vicki S., Jeffrey C. Valentine, and Harris M. Cooper. 2002. Interventions to increase physical activity among aging adults: a meta-analysis. *Annals of Behavioral Medicine: a Publication of the Society of Behavioral Medicine* 24(3):190–200.

Conner, Mark, and Christopher J. Armitage. 1998. Extending the Theory of Planned Behavior: A review and avenues for further research. *Journal of Applied Social Psychology* 28(15):1429–1464.

Consolvo, Sunny, David W. Mcdonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony Lamarca, Louis Legrand, Ryan Libby, Ian Smith, and James A. Landay. 2008. Activity sensing in the wild: A field trial of UbiFit Garden. In *Chi 2008 proceedings*, 1797–1806.

Eagle, Nathan, and Alex (Sandy) Pentland. 2006. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing* 10(4):255–268.

Eagle, Nathan, Alex Sandy Pentland, and David Lazer. 2009. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America* 106(36):15274–8.

Eston, Roger G., Ann V. Rowlands, and David K. Ingledew. 1998. Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children 's activities. *Journal of Applied Physiology* 84:362–371.

Farley, John U., Donald R. Lehmann, and Michael J. Ryan. 1981. Generalizing from "imperfect" replication. *Journal of Business* 54(4):597–610.

Finlay, Krystina A., David Trafimow, and Donna Jones. 1997. Predicting Health Behaviors From Attitudes and Subjective Norms: Between-Subjects and Within-Subjects Analyses. *Journal of Applied Social Psychology* 27(22):2015–2031.

Fishbein, Martin, and Icek Ajzen. 1975. *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, Massachusetts: Addison-Wesley Publishing Co.

Fjeldsoe, Brianna S., Alison L. Marshall, and Yvette D. Miller. 2009. Behavior Change Interventions Delivered by Mobile Telephone Short-Message Service. *American Journal of Preventive Medicine* 36(2):165–173.

Fogg, BJ, and Dean Eckles, eds. 2007. *Mobile Persuasion: 20 Perspectives on the Future of Behavior Change.* Palo Alto, California: Stanford Captology Media.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius. 2008. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research* 35(3):472–482.

Gollwitzer, Peter M., and John A. Bargh, eds. 1996. *The Psychology of Action: Linking Cognition and Motivation to Behavior.* New York, NY: The Guilford Press.

Hedtke, Paul. 2007. Personal Health Assistant in the Palm of Your Hand. In *Mobile persuasion: 20 perspectives on the future of behavior change*, ed. BJ Fogg and Dean Eckles, 71–76. Palo Alto, California: Stanford Captology Media.

Hendelman, D., K. Miller, C. Baggett, E. Debold, and P. Freedson. 2000. Assessment of moderate intensity physical activity in the field. *Medicine & Science in Sports & Exercise* 32(9):S442–S449.

Hogg, Michael A., Deborah J. Terry, and Katherine M. White. 1995. A Tale of Two Theories: A Critical Comparison of Identity Theory with Social Identity Theory. *Social Psychology Quarterly* 58(4):255.

Hughes, Nick, and Susie Lonie. 2007. M-PESA: Mobile money for the "Unbanked" Turning Cellphones into 24-Hour Tellers in Kenya. *Innovations: Technology, Governance, Globalization* 2(1-2):63–81.

International Telecommunications Union. 2010. Information and Communication Technology Statistics.

Janz, Kathleen F., John Witt, and Larry T. Mahoney. 1995. The stability of children's physical activity as measured by accelerometry and self-report. *Medicine & Science in Sports & Exercise* 27(9):1326–1332.

Kallgren, Carl A., Raymond R. Reno, and Robert B. Cialdini. 2000. A Focus Theory of Normative Conduct: When Norms Do and Do not Affect Behavior. *Personality and Social Psychology Bulletin* 26(8):1002–1012.

Kaplan, Warren A. 2006. Can the ubiquitous power of mobile phones be used to improve health outcomes in developing countries? *Globalization and health* 2:9.

Kass, Alex. 2007. Transforming the Mobile Phone into a Personal Performance Coach. In *Mobile persuasion: 20 perspectives on the future of behavior change*, ed. BJ Fogg and Dean Eckles, 63–69. Palo Alto, California: Stanford Captology Media.

Laforet, Sylvie, and Xiaoyan Li. 2005. Consumers' attitudes towards online and mobile banking in China. *International Journal of Bank Marketing* 23(5):362–380.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. Computational social science. *Science (New York, N.Y.)* 323(5915):721–3.

Liska, Allen E. 1984. A critical examination of the causal structure of the Fishbein/Ajzen attitude-behavior model. *Social Psychology Quarterly* 47:61–74.

Mallat, Niina, Matti Rossi, and Virpi Kristiina Tuunainen. 2004. Mobile banking services. *Communications of the ACM* 47(5):42–46.

Mathie, Merryn J., Adelle C. F. Coster, Nigel H. Lovell, and Branko G. Celler. 2004. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological Measurement* 25(2):R1–R20.

Miluzzo, Emiliano, Nicholas D. Lane, Kristóf Fodor, Ronald Peterson, Hong Lu, Mirco Musolesi, Shane B Eisenman, Xiao Zheng, and Andrew T Campbell. 2008. Sensing meets mobile social networks: The design, implementation and evaluation of the CenceMe Application. In *Sensys 2008*.

Miniard, Paul W., and Joel B. Cohen. 1981. An examination of the Fishbein-Ajzen behavioral-intentions model's concepts and measures. *Journal of Experimental Social Psychology* 17:309–329.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer. 2008. Coarse thinking and persuasion. *Quarterly Journal of Economics* 123(2):577–619.

Nolan, Jessica M., P. Wesley Schultz, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2008. Normative social influence is underdetected. *Personality and social psychology bulletin* 34(7):913–23.

Posner, Richard A. 1997. Social norms and the law: An economic approach. *The American Economic Review* 87(2):365–369.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. The constructive, destructive, and reconstructive power of social norms. *Psychological science : a journal of the American Psychological Society / APS* 18(5): 429–34.

Smith, Aaron. 2010. Americans and their gadgets. Pew Research Center, Pew Research Center Report.

Sohn, Timothy, Alex Varshavsky, Anthony Lamarca, Mike Y Chen, Tanzeem Choudhury, Ian Smith, Sunny Consolvo, Jeffrey Hightower, William G Griswold, and Eyal De Lara. 2006. Mobility detection using everyday GSM traces. In *Ubicomp lncs*, ed. P. Dourish and A. Friday, 212 – 224. Berlin, Germany: Springer-Verlag.

Strack, Fritz, and Roland Deutsch. 2004. Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review* 8(3):220–247.

Sunstein, Cass R. 1995. Social norms and social rules. *John M. Olin Law and Economics Working Paper Series* No. 36.

Terry, Deborah J., and Michael A. Hogg. 1996. Group norms and the attitude-behavior relationship: A role for group identification. *Personality and Social Psychology Bulletin* 22(8):776–793.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Trafimow, David, and Krystina A. Finlay. 1996. The importance of subjective norms for a minority of people: between subjects and within-subjects analyses. *Personality and Social Psychology Bulletin* 22(8):820–828.

Troiano, Richard P., David Berrigan, Kevin W. Dodd, Louise C. Masse, Timothy Tilert, and Margaret McDowell. 2008. Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise* 40(1):181–188.

Tufano, James T., and Bryant T. Karras. 2005. Mobile eHealth interventions for obesity: a timely opportunity to leverage convergence trends. *Journal of medical Internet research* 7(5):e58.

Verplanken, Bas, and Wendy Wood. 2006. Interventions to break and create consumer habits. *Journal of Public Policy and Marketing* 25(1):90–103.

Wantland, Dean J, Carmen J Portillo, William L Holzemer, Rob Slaughter, and Eva M McGhee. 2004. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. *Journal of medical Internet research* 6(4):e40.

Wechsler, Henry, Toben F. Nelson, Jae Eun Lee, Mark Seibring, Catherine Lewis, and Richard P. Keeling. 2003. Perception and reality: A national evaluation of social norms marketing interventions to reduce college students' heavy alcohol use. *Journal of Studies on Alcohol* 64:484–494.

Wood, Wendy. 2000. Attitude change: Persuasion and social influence. *Annual review of psychology* 51(1):539–570.

World Health Organization. 2003. South Africa: a novel approach to improving adherence to TB treatment. *Essential Drugs Monitor* 33:8.

Ybarra, Oscar, and David Trafimow. 1998. How priming the private self or collective self affects the relative weights of attitudes and subjective norms. *Personality and Social Psychology Bulletin* 24(4):362–370.