“Peer Pressure” for Good: Studying the Effect of Social Commitment on Task Completion

Jennifer Podracky, April Kim, Saurav Datta

Introduction

People spend each day completing a series of tasks; some of those tasks are ones they choose to do, but many of those are tasks that they do not want to but *have* to do. These tasks can be job responsibilities, chores, exercise, responsibilities related to taking care of a home or family, and so on.

What makes a person decide to complete a task today, instead of tomorrow? What is the key factor that drives a person from procrastination to task completion? Given humans' communal nature, one theory posits that telling someone else about their task drives action. By telling someone else that they're going to do something, an individual is making a social commitment to that other person; the social obligation to meet that commitment is what possibly drives them to actually complete the task. Over the course of this paper, we will explore the interaction between making a social commitment to others and completing the actual commitment.

Research Questions and Hypotheses

By conducting this experiment, we hoped to answer two causal questions related to this theory of social commitment. Firstly, does telling others about a task make a person more likely to complete that task? Secondly, does telling others in person about a task make them more likely to complete the task than if they told others via digital means (e.g. email, instant message, or test message)?

The goal of this study is to examine and test the following two research hypotheses:

- Telling other individuals about a task they intend to complete thereby makes them more likely to complete that task.
- Additionally, telling other individuals in person about a task makes them more likely to complete the task than telling other individuals via digital methods, e.g. text message, instant message, or email.

Signpost

Over the course of this report, we will describe the background and existing literature behind the underlying theory that social commitments of some form can lead to desired outcomes. We will also describe, in detail, the steps that we took to conduct our own experiment to determine the effect of social commitment on task completion in a specific scenario; this will include descriptions of the treatments, outcomes, subjects, randomization techniques and verification, and underlying assumptions. Finally, we will also analyze the results of the study that we conducted, discuss the implications of those results, and then examine issues and limitations we found with our study that could be addressed by future research.

Background

It has been observed that procrastinators keep putting off things unless there is a fire to put out. It is then that they become most focused, and take the task to completion[1]. The firefighting mode is triggered when a flight is about to be missed, a submission deadline is about to expire, a financial loss is about to happen, for example. To counter procrastination, the trick is to find the influence that triggers the firefighting mode. One such influence is public commitment.

Leveraging public commitment involves sharing your goals with people whose opinion matter to you, and who you would not want to disappoint[2]. These people maybe a partner, a friend, a colleague or followers on social media. Studies have shown that good intent by itself is not sufficient to complete a task[3]. Communicating publicly about your intent to meet your goals acts as a powerful incentive to actually taking an action to complete the task. Public commitment works because we crave an improved standing in society - we care what other people think of us, even if they are strangers[4]. Setting prior expectations creates a pressure forcing us to meet our goals.

Public commitment has the additional effect of creating a certain self-view[5]. Communicating about our intent to be regular at the gym may create a self-view that we are health conscious, for instance. This self-view in turn makes us behave consistently to meet our goals. Research terms such behavior as accountability. Being accountable to an audience creates the expectation of having to explain ourselves in case of a failure and makes us more likely to complete our goals.

Identification Strategy

The outcome that we need to measure to test these research hypotheses is whether or not our subjects completed a task that they intended to complete. In order to structure this in such a way that the task in question was of equal importance to all of our subjects, we assigned them a task as part of the study design; this ensured that everyone was working towards the same goal. This also allowed us to choose a task that both had a non-zero cost and was possible for us to verify after the fact. The task that we chose to assign to our subjects was to walk at least 5,000 steps over the course of a 24 hour period, and we asked them to verify using a screenshot of a step tracking app or device.

Key assumptions that were made to attempt to make a causal inference from this experiment were excludability and non-interference.

The exclusion restriction assumption assumes that the potential outcome of a subject responds only to the treatment that they received, and that their treatment assignment is irrelevant. In this experiment, this means that whether or not a person walked 5,000 steps was only affected by whether or not they told others about their assigned task, and *not* affected by the email that the authors sent them containing their treatment assignment. Given the fact that the majority of our subjects personally knew at least one of the authors, we could argue that this assumption is not valid due to the surveyor demand affect. We will discuss this further in the Limitations section of this paper.

The non-interference assumption states that the potential outcome for a subject is dependent only on whether or not that subject received the treatment, and is independent of the treatment that other subjects received. In this study, this specifically means that whether or not a subject completes their number of steps is independent of what treatment other subjects received. Because of the social nature of this experiment, and because many of the subjects knew at least one of their fellow subjects, the authors had concerns about this assumption from the inception of this experiment. We attempted to mitigate the risk of violating this assumption by asking subjects not to communicate with other subjects about their task or treatment assignment; the majority of subjects knew who else was participating in the study with them, so we assume that this guideline was followed. We also ensured that every subject was assigned the same task, so even though their treatment assignments may have been different, they all had the same potential outcomes (task completion or non-completion). However, it is possible that this non-interference assumption was violated; if three subjects lived together (which was a possible scenario) and were each assigned to a different treatment level, one telling another in person about their task may have encouraged the other two to complete their own task. Ideally, though, this violation would not have occurred if the subjects followed the stated guidelines about communicating with other subjects.

Experimental Design

To test the research hypotheses stated in the introduction, we conducted two studies (one pilot and one final) using 93 total participants to look specifically at how making a commitment to other people made an individual more or less likely to complete an assigned task of walking at least 5,000 steps in a 24 hour period. Additionally, we looked at the difference in effect between communication methods (in-person versus digitally). We created a within-subject comparison by varying the methods of treatment between the same individuals over multiple days.

The pilot study took place over the week of November 26-30, 2018. The 18 pilot subjects were randomly assigned to their treatment level assignments on Monday, November 26. Treatment assignments were sent out that Monday evening via email. These assignments contained a description of the task that the individual was asked to complete each day for the next three days (Tuesday through Thursday). The assignments also contained a description of the communication method that the subject was to use to tell others about their task each day. A data collection survey was then sent out to all pilot subjects on Friday evening, for subjects to return with their outcomes from the previous three days.

The full study took place over the week of December 3-7, 2018. The 75 full study subjects were randomly assigned to their treatment level assignments on Monday, December 3. Treatment assignments were sent out that Monday evening via email. These assignments contained a description of the task that the individual was asked to complete each day for the next three days (Tuesday through Thursday). The assignments also contained a description of the communication method that the subject was to use to tell others about their task each day. A data collection survey was then sent out to all full study subjects on Thursday evening, for subjects to return with their outcomes from the previous three days. See Appendix for the treatment assignment delivery messaging and the data collection survey.

Subjects

Subjects for this experiment were primarily collected using the authors' social networks. Subjects were also collected by recruiting fellow MIDS students and MIDS faculty. Finally, some subjects were recruited using Craigslist listings requesting volunteers. Potential participants were incentivized to participate by offering the chance to be entered in a raffle for $25 Amazon gift cards upon successful submission of their final data. Twenty gift cards were available, and each subject would only be eligible to receive one gift card maximum.

Potential participants were asked a set of demographics questions to establish their general age, gender, and location. They were asked if they had access to a device that would enable them to communicate with others digitally, e.g. via text message, instant message, or email. (All potential participants answered yes to this question). Finally, they were also asked if they had access to an app, smartphone, or other device that would enable them to track their steps over a 24-hour period, and if they would be able and willing to share a screenshot or other image showing the number of steps they had taken (see Appendix). As long as the participant was at least 18 years old, answered in the affirmative to the last two questions (confirming that they had a device to communicate digitally, and could track and share their step count), and had a valid email, then they were included as a subject in our study. Of the 150 responses we received from potential participants, we collected 93 eligible subjects.

Of the 93 total subjects, 18 were randomly chosen to participate in a pilot study conducted the week prior to the full study. The final study was completed with the 75 remaining participants.

Covariates

During recruitment, we asked each subject to (optionally) report their gender, general age range, home city, and whether or not they lived with others. The possible levels of each covariate that we collected, as well as the encoding used for each level, are listed in the Appendix.

The full text of the recruitment survey, including the questions we asked to gather these covariates, can be found in the Appendix.

Treatment

The treatment in this experiment is whether or not a subject notifies other non-subjects about their intent to complete the task. There are three levels of the treatment: the control in which the subject tells no one else, one treatment group in which the subject tells others via digital communication methods, and a second treatment group in which the subject tells others in person.

Each subject in both the pilot and full studies was assigned to each level of the treatment over three days, with a different treatment each day. To control for timing effects, each subject was randomly assigned to one of six treatment assignment sequences (which cover the six possible ways that the three levels of treatment could be ordered). The treatment assignments are defined below using ROXO grammar, and the following encodings and treatment definition symbols:

ROXO Encoding	**Treatment**
-	(Control) Notify no one else about the assigned task
X1	Notify at least 2 people about the assigned task via digital means
X2	Notify at least 2 people about the assigned task in person

Table 1. Treatment level encodings and definitions

Treatment Group	**Pre-Study**	**Day 1**		**Day 2**		**Day 3**	
1	R	-	O	X1	O	X2	O
2	R	-	O	X2	O	X1	O
3	R	X1	O	-	O	X2	O

4	R	X1	O	X2	O	-	O
5	R	X2	O	-	O	X1	O
6	R	X2	O	X1	O	-	O

Table 2. ROXO experimental design

Treatment assignments were delivered on Day 0 (the day prior to Day 1), via email. A sample treatment assignment description, and the full text of the treatment delivery email, can be found in the Appendix.

Randomization Engineering

During the pilot study, 18 participants were randomly assigned to one of six treatment sequences. The assignment was done such that 3 subjects were assigned to each treatment sequence. This was accomplished by creating a vector of length 18, with three instances of each possible treatment sequence (1 through 6). The treatment assignment vector was then generated by using `sample()` on this vector to randomly reorder it. This treatment vector was then applied to the pilot subjects.

During the full study, 75 participants were randomly assigned to one of six treatment sequences. The assignment was done such that odd-numbered treatment sequences had 12 subjects in each, and even-numbered sequences had 13 subjects in each. This was accomplished by creating a vector of length 75, with 12 instances of each possible odd treatment sequence (1, 3, and 5) and 13 instances of each possible even treatment sequence (2, 4, and 6). The treatment assignment vector was then generated by using `sample()` on this vector to randomly reorder it. This treatment vector was then applied to the final study subjects.

Covariate Balance Check - Treatment Assignment

Using logistic regression, we checked to see if any of our covariates are statistically significant predictors of our treatment assignment. We used logistic regression rather than linear regression because each treatment sequence, although assigned an integer id, is not numeric or linear. If our randomization was successful, none of the covariates should predict the treatment in the generated regression model. The summary of the logistic regression of our covariates on the treatment assignment is below.

	Dependent variable:
	Treatment sequence
Female	0.494 (0.720)
Gender non-conforming	17.753 (6,522.639)
Ages 25-34	−1.108 (1.147)
Ages 35-44	−0.949 (1.604)
Ages 45-54	−0.677 (1.567)
Ages 55-64	15.752 (6,522.639)
Ages 65+	16.208 (4,588.306)
Agest other	15.725 (6,522.639)
Has housemates	−16.687 (2,542.957)
Housemates unknown	−0.898 (7,000.817)
Knows authors	−0.822 (1.215)
Knows authors unkonwn	0.671 (1.626)
Latitude	−0.191 (0.182)
Longitutde	0.024 (0.025)
Constant	29.095 (2,542.973)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3. Logistic regression model coefficient estimates, predicting treatment sequence using the available covariates.

None of our covariates are statistically significant predictors for our treatment assignment, again verifying that our randomization methods were successful.

Outcomes

We observed two outcomes from each day that the subject reported day. First, we recorded a binary variable indicating whether the subject walked at least 5,000 steps in a day, or not. Second, we recorded the number of steps that the subject walked in the day (regardless of whether they had reached 5,000 steps or not).

Each subject's daily outcomes were collected at the end of the study, via a data collection survey sent after study completion (see Appendix). For those subjects who completed it, the authors asked for each day: "How many steps did you take on [this day]?"

The authors also asked for screenshots or other device images showing the number of steps the subject recorded in their response. This was done to verify that subjects were not falsely stating that they had completed the task when they had not done so.

The authors collected outcome data from this response survey, and do not have data on outcomes for those subjects who did not complete the data collection survey.

Compliance Check, Non-Compliance, and Attrition

To ensure that our subjects actually completed the treatments that were assigned to them, we conducted a compliance check as part of the data collection survey sent after study completion (which was also used to collect outcomes; see Appendix). The survey asked two compliance-related questions for each day:

- How many people did you tell about the task?
- What method did you use to complete your communication: digital, in-person, or a mix?

This is not a foolproof compliance check, as subjects may have falsely reported which treatment they completed in order to align with the treatments they were assigned; however, we emphasized at multiple points during the study that truthfulness was paramount, so we hope that this mitigated any subjects' inclination to lie. The authors collected compliance data from this response survey, and matched the subject's reported treatment to their assigned treatment for that day.

The first compliance check was meant to verify that subjects told as many people as we asked them to (specifically, either no one or at least two other people). However, we found that multiple subjects who complied otherwise failed in a different way by only telling one other person on one of their three days of observation. We evaluated potentially removing all 14 subjects from our compliant dataset. However, after talking to some of the subjects in question, it became clear that there was some confusion about when the subjects were supposed to tell others about their task, and if they had to stop telling others once they hit their step goal. In one instance, the subject reached 5,000 steps before he had notified the second person; he then decided not to tell them at all, because he thought he'd already failed the given task. Because of this confusion, and because the difference between telling one or two people was not an effect we were trying to capture in this study, we considered these subjects to be compliant in our analysis as long as they told at least one other person.

The second compliance check attempts to verify with the subject that they performed the communication method assigned to them, and didn't self-select into one of the other communication methods. If the subject's reported communication method did not match their assigned treatment, e.g. the subject reported that they told no one when they were assigned to tell others in person, then the subject was determined to be non-compliant.

Because treatment compliance and outcomes were both measured at the end of the study, the authors do not have data on whether or not the subjects completed their assigned treatments (or other treatments) for those subjects who did not complete the data collection survey.

We also experienced attrition. Of our 75 final study subjects, 51 subjects returned completion data that contained both their compliance and outcome data. For the other 24 subjects (32%), we received no response and have no data on their compliance or outcomes for any of the three days.

To determine if the rate of attrition was relatively random across the treatment sequences, we performed a linear regression using the treatment sequences and the available covariates to predict whether a subject submitted data to us, or if they attrited. These results are below:

	Dependent variable:
	Final survey submitted
Treatment Seq 2	−0.047 (0.156)
Treatment Seq 3	−0.013 (0.150)
Treatment Seq 4	−0.054 (0.149)
Treatment Seq 5	−0.240 (0.149)
Treatment Seq 6	−0.240* (0.143)
Ages 25-34	0.153 (0.117)
Ages 35-44	0.168 (0.170)
Ages 45-54	0.405** (0.176)
Ages 55-64	0.327 (0.372)
Ages 65+	0.265 (0.281)
Agest other	−0.277 (0.392)
Female	−0.023 (0.087)
Gender non-conforming	−0.244 (0.399)
Has housemates	0.153 (0.158)
Housemates unknown	0.472 (0.395)
Knows authors	0.278** (0.135)
Knows authors unkonwn	−0.615*** (0.163)
Latitude	0.041** (0.019)
Longitutde	−0.004 (0.003)
Constant	−1.604 (0.995)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4. OLS regression coefficient estimates, predicting study attrition from treatment assignment and covariates.

Neither treatment sequence nor our covariates were statistically significant predictors of submission/attrition, which leads us to believe that attrition was relatively randomly distributed across the treatment groups. There is a marginally significant estimate for the Treatment Sequence 6 indicator; however, due to the high number of covariates that we're applying in this model (17 in total), it was likely that one of them would be a significant predictor just by chance.

Pilot Study Learnings

Of the 93 total subjects that were recruited for this experiment, 18 were randomly chosen to participate in a pilot study conducted the week prior to the full study. Treatment assignments were sent out on Monday evening, subjects were asked to complete tasks Tuesday through Thursday, and then the data collection survey was sent out to all pilot subjects on Friday evening.

By running the pilot, we identified several issues with our delivery of the treatment assignments. We received several questions after the treatment was sent out on Monday evening, asking to clarify who subjects could tell and how much they were to supposed to tell others about the experiment. We modified our treatment assignment communication to address those questions as part of the original email, and received many fewer questions during the full study. We also modified the initial assignment email to remind our subjects to turn on their step trackers; 2 potential subjects in the pilot contacted me to say they had forgotten to turn their devices on. We still had subjects in the final study that didn't comply for this reason, so that communication wasn't as effective.

We also identified an issue with the delivery of our follow-up data collection survey, which we depended on to actually collect treatment compliance and outcome data. The majority of our subjects were based on the East Coast; however, because of a scheduling issue, the data collection survey was sent out to subjects at 9pm EST on Friday. By Saturday evening, we had a 33% response rate (6 subjects), which was lower than we hoped to have. We followed up on Sunday and received no additional responses. The feedback that we got by reaching out to pilot participants outside of the constraints of the study was that they didn't check their emails over the weekend, so both of our collection emails missed them. To address this, we changed our strategy in the final study to send the initial collection email on Thursday night, sent a follow-up on the following Monday morning, and also added a requirement to submit by midnight on Monday in order to qualify for the gift card drawing. After making these changes, our response rate for the final study was 68% (more than double our pilot response rate).

Finally, while analyzing our pilot study results, it also occurred to us that there may be a difference in potential outcomes between those subjects that knew one of the authors personally (e.g. a friend, spouse, or family member) versus those subjects that were recruited via Craigslist and therefore had no personal relationship with any of the authors. We hypothesized that there may be some form of social commitment to us as the authors, even if no communication was made to other people, for our friends and family that we'd recruited to participate. To make sure we captured this variable, we included an additional question asking "Do you personally know one of the individuals conducting the study?" in the final study data collection survey. We could have generated this data ourselves, but by asking the subjects themselves, we got their own impression of their relationship with the authors (instead of the authors' perception of their relationship with the subject).

Results

Evaluating Raw Data

We conducted a thorough evaluation (in addition to the covariate balance check) and data cleaning prior to estimating the regression models. The key takeaways were:

- Two subjects had reported communicating both in person and through digital means on a day when they were assigned treatment. We marked these subjects as compliant by changing "both communication methods" to a communication method different from what they had reported as already having performed. This specific interpolation was done as one of our hypothesis is testing the difference between treatment and no treatment. Since these subjects had reported as having been treated, albeit one additional level than what was assigned, we marked them as having complied with the assigned treatment sequence.

- 19 out of 49 subjects who reported communication methods and step counts had performed communication methods in a different order from the one assigned to them and marked as non-compliant (see Compliance Check, Non-Compliance, and Attrition).
- Of the 19 non-compliant subjects, four subjects performed three different treatments over the span of three days. We treated these subjects as if they were assigned to a treatment sequence matching the order in which they had communicated (see Sequencing Effect).
- We further validated through checking for carryover effect (see Sequencing Effect) and covariate balance check (see Covariate Balance Check - Submitted Data) that adding the four unique cases did not significantly change the distribution of covariates across our treatments.

Sequencing Effect

Although our within-subject design exposed each subject to both communication methods (in-person and digital), we were interested in understanding the effect of different types of social commitment, which would require reformatting our data in similar format as that of between-subjects design. In order to rule out any sequencing effect, we applied linear regression (Equation 1) to test if the treatments assigned on first two days would be highly predictive and significant of step count on Day 3.

$$y_i = \beta_0 \ + \ \beta_1(Treatment\ on\ Day\ 1) + \beta_2(Treatment\ on\ Day\ 1) + \varepsilon_i$$

Equation 1. Short linear model for sequencing effect.

In addition, we fitted a second model by including each subject’s step counts from first two days as covariates (Equation 2) as to understand the subject’s step count change as a function of treatment against how many steps they would typically take on any given day.

$$y_i = \beta_0 \ + \ \beta_1(Treatment\ on\ Day\ 1) + \beta_2(Treatment\ on\ Day\ 1) + \\ \beta_3(Step\ Count\ on\ Day\ 1) + \beta_4(Step\ Count\ on\ Day\ 2) + \varepsilon_i$$

Equation 2. Long linear model with step counts as covariates for sequencing effect.

Table 5 shows the regression results from two models to identify any potential sequencing effect. In the first model, we would expect that subjects who made a digital commitment on Day 1 would take on average 228 less steps on Day 3, and subjects who made an in-person social commitment on Day 1 would take on average 1,105 less steps than those who made no social commitment, holding all other variables constant. Observing the treatment variables on Day 2, subjects who made a digital commitment would take on average 272 less steps on Day 3 and subjects who made an in-person commitment would take 779 less steps compared to those who made no social commitment, holding all other variables constant. No significant treatment effect was detected based on the regression coefficients of treatment assignment. Similarly, the second regression model did not indicate any significant treatment effect using the same regression coefficients. Two covariates (number of steps taken on Day 1 and Day 2) were significantly associated at $P<0.1$ and $P<0.01$, respectively, with higher step counts on Day 3.

	Dependent variable:	
	Steps - Day 3	
	Short	Long
	(1)	(2)
Digital - Day 1	−228.177 (1,243.893)	−495.009 (948.403)
In person - Day 1	−1,104.705 (1,560.442)	−841.643 (1,188.417)
Digital - Day 2	−272.684 (1,452.143)	−857.124 (1,110.653)
In person - Day 2	−778.868 (1,242.131)	−560.681 (951.815)
Steps - Day 1		0.254* (0.127)
Steps - Day 2		0.453*** (0.137)
Constant	7,565.343*** (1,065.706)	2,681.342** (1,180.560)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 5. Short and long regression coefficient estimates on the number of steps taken on Day 3.

No significant treatment effect was found in both models and we ruled out the possibility of sequencing effect.

Data Transformation for Analysis

The day indicators were collapsed to have one treatment variable D (maintaining three levels of treatment to test the hypothesis that in-person commitment makes an individual more likely to complete a task than making a digital commitment would) and single step count variable. To test the hypothesis that telling others about a task would make one more likely to complete a task, we created a second treatment indicator D_2 where digital and in-person commitment were combined as “making a social commitment” (denoted by no commitment = 0, social commitment = 1).

Covariate Balance Check - Submitted Data

We performed two separate tests to ensure covariate balance on the re-formatted data. Observing the overall population of each categorical covariate (Figure 1), there were clear imbalance in age, indicator for having a housemate as well as indicator for whether the subject knows one of the authors. Using logistic regression, we fitted two separate models using D (three levels of treatment) and D_2 (two levels of treatment) as the treatment variable to ensure our covariates are still not significantly predictive of the treatment assignment and balance is maintained (Table 6).

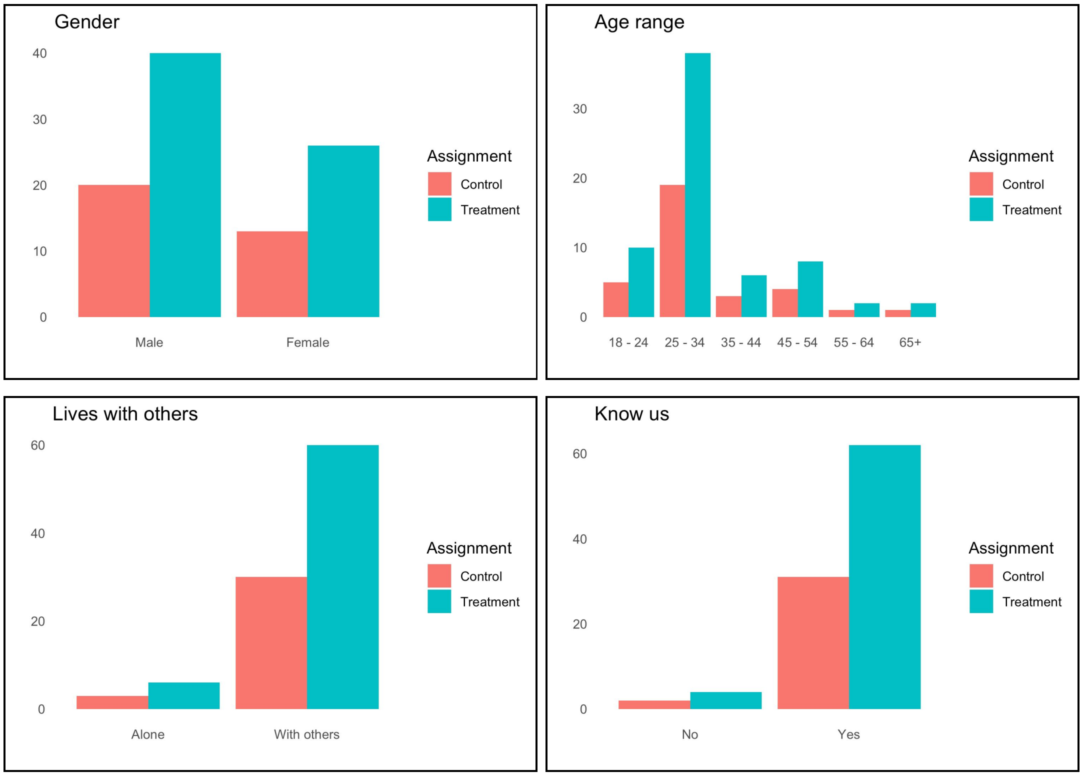

Figure 1. Overall population of four different categorical variables.

	Dependent variable:	
	2 levels treatment	3 levels treatment
	(1)	(2)
Female	0.000 (0.494)	0.000 (0.494)
Ages 25-34	−0.000 (0.660)	−0.000 (0.660)
Ages 35-44	0.000 (0.977)	0.000 (0.977)
Ages 45-54	0.000 (0.876)	0.000 (0.876)
Ages 55-64	0.000 (1.406)	0.000 (1.406)
Ages 65+	0.000 (1.402)	0.000 (1.402)
Has housemate	0.000 (0.835)	0.000 (0.835)
Knows us	0.000 (1.054)	0.000 (1.054)
Latitute	0.000 (0.102)	0.000 (0.102)
Longitude	−0.000 (0.016)	−0.000 (0.016)
Constant	0.693 (5.308)	0.693 (5.308)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 6. Logistic regression model coefficient estimates, predicting three levels of treatment assignment and two levels of treatment assignment using the available covariates.

None of our covariates are statistically significant predictors for the two different treatment assignments, verifying that balance is maintained from before collapsing.

Results Visualizations

Observing the distributions of steps taken in different levels of treatment, we see that median step count in no social commitment group is less than those of digital commitment and in-person commitment, as well as the two modes of commitment combined (Figure 2).

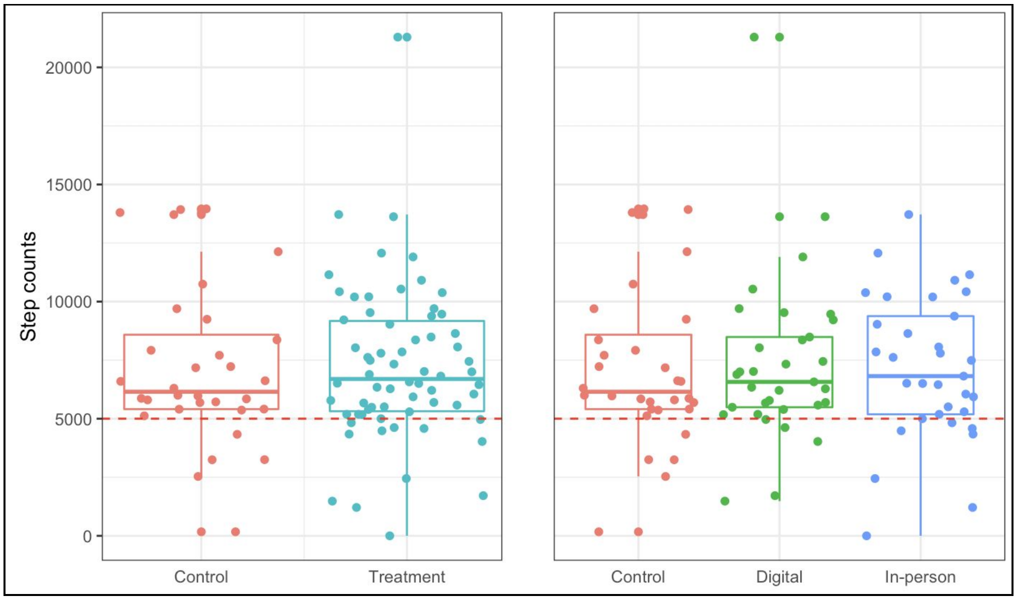

Figure 2. Visual comparisons of step count distributions from two levels of treatment assignment on the left, and three levels of treatment assignment on the right. Red dashed line indicates threshold equivalent to 5,000 steps.

Regressions

To test the hypothesis that making a social commitment about a task they intend to complete would make an individual more likely to complete that task, we fitted two regression models using two levels of treatment and subject ID in the primary model (Equation 3) and covariates in the secondary model (Equation 4). We included subject ID, which is an individual level indicator, in the primary model to pull off the control (no commitment) average for each subject while the treatment indicator would give the average difference between control and treatment (social commitment). The standard errors were clustered based on subject ID. We fitted a secondary model using covariates over subject ID with the expectation that covariates would increase the efficiency of treatment estimate with higher standard error compared to that of the primary model. Robust standard errors were calculated for the secondary model.

$$y_i = \beta_0 + \beta_1(3\ Level\ Treatment) + \beta_2(Subject\ ID) + \varepsilon_i$$

Equation 3. Primary linear regression model using three levels of treatment, predicting if greater than 5,000 steps were taken using treatment assignment and subject ID.

$$y_i = \beta_0 + \beta_1(3\ Level\ Treatment) + \beta_2(Age\ Range) + \beta_3(Gender) + \beta_4(Has\ Housemates) + \beta_5(Knows\ Authors) + \beta_6(Latitude) + \beta_7(Longitude) + \varepsilon_i$$

Equation 4. Secondary linear regression model using three levels of treatment, predicting if greater than 5,000 steps were taken using treatment assignment and covariates.

Comparing the two models (Table 7), no significant treatment effect was found in both primary and secondary models, with similar average treatment effects. As expected the standard error of treatment indicator is bigger in the secondary model compared to that of the primary model, indicating that subject ID is more predictive of the outcome than gender, general age range, home city, and whether or not they live with others and know us are. Although not shown in Table 7, majority of individual level indicators is positively significant at $p<0.01$ level and highly predictive of whether the subject took more than 5,000 steps regardless of the treatment assignment, which is in line with bigger standard error of treatment indicator in the secondary model (full table in Appendix Table 10). Two oldest age groups show significant treatment effect at $p \leq 0.05$ level in the secondary model and as we had hypothesized after the pilot study, whether or not the subjects know the authors is positively associated with the outcome variable at $p<0.1$ level.

We fail to reject the null hypothesis that telling other individuals about a task they intend to complete is not any more or less likely to make them complete that task.

	Dependent variable:	
	Steps > 5000	
	User ID	Covariates
	(1)	(2)
Social commitment	−0.047 (0.064)	−0.043 (0.074)
Ages 25-34		0.157 (0.140)
Ages 35-44		0.156 (0.140)
Ages 45-54		0.086 (0.160)
Ages 55-64		0.369*** (0.143)
Ages 65+		0.340** (0.144)
Female		−0.068 (0.076)
Has housemate		0.303* (0.155)
Knows us		0.237* (0.131)
Latitute		0.030 (0.019)
Longitude		−0.004 (0.003)
Constant	0.031 (0.045)	−1.315 (0.970)
Subject IDs ommitted	Yes	No
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 7. Primary and secondary linear regression coefficient estimates using two levels of treatment.

To test the second hypothesis that making an in-person social commitment about a task would make a subject more likely to complete the task than making a digital social commitment would, we fitted our primary model using three levels of treatment assignment and subject ID (Equation 5) and covariates in the secondary model (Equation 6) with similar assumptions and expectations as described for testing of the first hypothesis.

$$y_i = \beta_0 + \beta_1(2\ Level\ Treatment) + \beta_2(Subject\ ID) + \varepsilon_i$$

Equation 5. Primary linear regression model using three levels of treatment, predicting if greater than 5,000 steps were taken using treatment assignment and subject ID.

$$y_i = \beta_0 + \beta_1(2\ Level\ Treatment) + \beta_2(Age\ Range) + \beta_3(Gender) + \beta_4(Has\ Housemates) + \beta_5(Knows\ Authors) + \beta_6(Latitude) + \beta_7(Longitude) + \varepsilon_i$$

Equation 6. Secondary linear regression model using two levels of treatment, predicting if more than 5,000 steps were taken using treatment assignment and covariates.

In both regression results (Table 8) no significant treatment effect on taking more than 5,000 steps is found when comparing no commitment to making a digital commitment and separately to making an in-person commitment. We found majority of subject ID level indicators to be highly predictive of whether the subject took more than 5,000 steps regardless of the treatment assignment at $p<0.01$ level in the primary model (full table in Appendix Table 11). Two oldest age groups show significant treatment effect at $p<0.05$ level and whether or not the subjects know the authors is positively associated with more than 5,000 steps at $p<0.1$ level in the secondary model.

	Dependent variable:	
	Steps > 5000	
	User ID	Covariates
	(1)	(2)
Digital commitment	−0.001 (0.065)	0.003 (0.078)
In person commitment	−0.092 (0.076)	−0.088 (0.092)
Ages 25-34		0.157 (0.139)
Ages 35-44		0.156 (0.141)
Ages 45-54		0.086 (0.157)
Ages 55-64		0.369** (0.144)
Ages 65+		0.340** (0.145)
Female		−0.068 (0.076)
Has housemate		0.303* (0.160)
Knows us		0.237* (0.132)
Latitute		0.030 (0.019)
Longitude		−0.004 (0.003)
Constant	0.031 (0.049)	−1.315 (0.951)
Subject IDs ommitted	Yes	No
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 8. Primary and secondary linear regression coefficient estimates using three levels of treatment.

We fail to reject the null hypothesis that there is no difference in effect between in-person commitment and digital commitment to make an individual more likely to complete a task.

Power Calculations

Since we fail to reject the null hypothesis that task completion followed by making a social commitment is any more or less likely to be completed, we performed power analysis to 1) understand how powerful our experiment is in detecting an effect of social commitment with a significance level of 0.05 and 2) determine the number of subjects required for a test with 80% power. We initially decided to use Hedges' *g* as a measure of effect size given the differences in two groups' outcome variable standard deviations (0.37, 0.40) and sample sizes in respective groups. With both groups of medium sample sizes (sample size > 20), however, we expected Hedges' *g* and Cohen's *d* to be roughly equivalent and both statistics to perform about the same. We found both measures to be equivalent as expected (Hedges' $g = 0.103$, Cohen's $D = 0.104$) and used Cohen's *D* as effect size for the power calculations.

Computing the power of two sample *t*-test, we have 7.72% power to detect a social commitment effect with a significance level of 0.05 (Figure 3). This result is largely due to a very small effect size (0.104) where the control and treatment groups' means only differ by 0.1 standard deviation. This indicates only trivial difference in the outcome variable of more than 5,000 steps in the two groups.

```
 t test power calculation

         n1 = 33
         n2 = 66
          d = 0.1042079
  sig.level = 0.05
      power = 0.07723521
alternative = two.sided
```

Figure 3. Power calculations based on 33 control subjects, 66 treatment subjects and Cohen's D effect size determined from whether or not subject has taken more than 5,000 steps, and 0.05 significance level.

To detect how many subjects we need to observe for a test with 80% power, we computed the power of one sample *t*-test using the same effect size as the previous power test (Figure 4). For desired power of 80% we would need about 1,450 subjects in each group for a total of 2,900 subjects. We, however, have an experiment that is of within-subject design and thus the total number of subjects needed can be "as little as" 970 subjects total if we continue to find no carryover effect and maintain covariate balance with more subject recruitment and data collection. To state that our study is underpowered is not an exaggeration as we had significantly less subjects in the current study.

```
 Two-sample t test power calculation

          n = 1446.518
          d = 0.1042079
  sig.level = 0.05
      power = 0.8
alternative = two.sided
```

Figure 4. Power calculations based on 80% power, Cohen's D effect size determined from whether or not subject has taken more than 5,000 steps, and 0.05 significance level.

Discussion

Limitations and Future Study

The generalizability of the results of this study are limited due to the subjects that participated in the study, as well as the task that we asked our subjects to complete and measured as our outcome. Our results will not generalize to the larger US population because of the heavy distribution of subjects in the 25-34 year old age group, nor will it generalize to just that age group because the subjects that we did recruit were primarily concentrated on the east and west coasts of the US. It is also difficult to generalize our findings to other tasks that aren't specifically walking, like completing household chores, running errands, or any other non-physical task.

One limitation with the way that we structured this study is that all of our data is self-reported. We relied on our subjects to complete the treatment that we had assigned to them, and then to accurately report back to us at the end of the week. This opened us up to human error; subjects may have falsified compliance out of a desire to please the authors. They also may have simply misremembered what they did on earlier days, given that we asked for Tuesday's treatment and outcomes on Friday. Ideally, future studies or further research on this hypothesis could be conducted in a more controlled setting that allows researchers to independently verify treatment compliance, without relying on the subject's own report.

Additionally, the way that the study was structured made it very difficult to separate out the effect of social commitments to others from the "surveyor demand effect", or the effect of us as the study authors asking us to complete the task. This was especially apparent within subjects that knew one of the authors personally; one coworker came up to their author contact on each day of the study to update them on their step count throughout the day.

Other enhancements for future studies on this topic would include more subjects, clustering by household to limit non-interference assumption violations, and adding additional covariates. More subjects in this study would have resulted in more power, and would have helped us find a significant difference even with our small treatment effect. Clustering by household would allow us to control for the average activity level within a household, and the effect of sharing households with others who generally tend to walk more on one's own walking habits. Additionally, gathering and utilizing more covariates that could predict the outcome of walking more, including method used to commute to work, average amount of exercise in a week, and any physical restrictions.

Conclusion

We set out to try to answer the question, "Does making a commitment to other people to do something then make you more likely to actually do it?" using a within-subjects, randomized experiment. We conceptualized this by assigning study subjects the task to walk 5,000 steps in a day, and to tell their friends and family about that task. We also varied whether or not we asked them to tell others in person or digitally. Among the subjects who responded and complied with our treatment assignments, we collected step count measurement and overall task completion status (yes/no). We did not find a statistically significant effect of social commitment on task completion, nor did we find a statistically significant difference in effect between in-person commitment and digital commitment. We failed to reject the null hypothesis and conclude that not enough evidence is available to suggest the null is false at the 95% confidence level. However, the authors had several learnings regarding designing and structuring causal experiments, and have several ideas for continued and improved research on this topic.

References

1. Sivori, C. (2008, August 26) Tips for working with procrastinators. *Letter Never Sent.* Retrieved from www.letterneversent.com/tips-for-working-with-procrastinators/2462/
2. Evans, K. (2016, July 12) How to Stop Procrastinating and Become More Productive. *Intelligent Change.* Retrieved from www.intelligentchange.com/blogs/news/how-to-stop-procrastinating-and-become-more-productive
3. Gollwitzer, P. M., & Sheeran, P. (2006). Implementation Intentions and Goal Achievement: A Meta‐analysis of Effects and Processes. *Advances in Experimental Social Psychology*, *38*, 69–119. doi.org/10.1016/S0065-2601(06)38002-1
4. Webb, C. (2016, July 29) How to Beat Procrastination. *Harvard Business Review.* Retrieved from hbr.org/2016/07/how-to-beat-procrastination
5. Gollwitzer, P. M., & Sheeran, P. (2006). Implementation Intentions and Goal Achievement: A Meta‐analysis of Effects and Processes. *Advances in Experimental Social Psychology*, *38*, 69–119. https://doi.org/10.1016/S0065-2601(06)38002-1

Appendix: Tables and Figures

Code and Files

Relevant data files and R Markdown files can be found at
https://github.com/jmercado14/mids-w241-social-commitment

Subject Recruitment Survey

Subjects for the pilot and full studies were recruited via multiple channels, and asked to sign up to be considered for the experiment by filling out a survey collecting basic demographic information. They were also asked if they had access to a device that would enable them to communicate with others digitally, e.g. via text message, instant message, or email. (All potential participants answered yes to this question). Finally, they were also asked if they had access to an app, smartphone, or other device that would enable them to track their steps over a 24-hour period, and if they would be able and willing to share a screenshot or other image showing the number of steps they had taken. This recruitment survey was structured as follows:

Question Purpose	***Question Text***
Informational	**Welcome!** Thank you for your interest in this study! We are a group of Berkeley MIDS students taking w241, and we are interested in learning if making a social commitment helps you stay committed to complete a task. If you choose to participate, you will be entered into a raffle to win a $25 Amazon gift card. **Requirements for participation:** You will be asked to complete a task, assigned by the researchers, after making a social commitment to do so. You may be given specific instructions on what communication method to use and how many people to communicate to. Lastly, you will also be asked to participate in a 5-minute follow-up survey within a few days of your task completion. If interested in participating, please fill out the brief form that follows. We'll be in touch! Jennifer Podracky, April Kim, and Saurav Datta (If you have any questions or concerns, please reach out to jennifer.m.podracky@berkeley.edu, mjapkim@berkeley.edu or sauravdatta@berkeley.edu)

Informational	**The following questions will be used to contact you. They will not be released or published in any form as a part of this study.**
Contact	What is your name?
Contact	What is the best email to contact you at?
Informational	**The following questions will be used to block your random assignment to the different treatment groups, and to see if treatment effects vary across different groups.** This information will **not be used in any other way**, and is **optional.**
Covariate	What is your gender? ● Male ● Female ● Gender non-conforming ● Other: [blank] ● Prefer not to answer
Covariate	What is your age? ● 18 - 24 ● 25 - 34 ● 35 - 44 ● 45 - 54 ● 55 - 64 ● 65+
Covariate	Where do you live (City, State if applicable, Country)?
Covariate	Do you live alone or with others? ● Alone ● With others
Informational	**The following questions will verify your ability to participate in this study.**
Participation Requirement	Do you have access to a device that can track the number of steps you take over the course of a 24 hour period, e.g. a smartphone/iPhone, wearable fitness tracker, or pedometer? If so, would you be willing to share that data with us if asked? ● Yes ● No
Participation Requirement	Do you have a device that allows you to text, email, or otherwise send a digital message to others? (e.g. cell phone, smart phone, laptop with wifi connection, etc.)

	• Yes • No

Please note that questions with bullets listed beneath them are multiple choice; the bullets list the possible answers that participants had to choose from. All questions other than the multiple choice questions accepted a freeform text response.

Treatment Assignment Delivery

Subjects who responded to the recruitment survey and then qualified for this experiment were sent the following message on the Monday prior to the beginning of the study:

> Hello,
>
> Thank you again for signing up to be a part of our Berkeley social commitment study! The study will take place over the next three days, from **Tuesday**, **December 4 to Thursday, December 6.**
>
> Your task for each of the next three days will be to **walk at least 5,000 steps over the course of each day.**
> On Tuesday, we ask that you tell **at least two people *in person* (and no one in digital form)** that you've been asked to walk at least 5,000 steps.
> On Wednesday, we ask that you tell **no one else** that you've been asked to walk at least 5,000 steps.
> On Thursday, we ask that you tell **at least two people *in digital form*, e.g. text, email, etc. (and no one in person)** that you've been asked to walk at least 5,000 steps.
>
> Please do not share the instructions you've been given regarding the method of communication you are to use, *especially* with other study participants, as the sharing of this information could introduce bias to our experiment. For example, you can say "For a study I'm participating in, I need to walk 5,000 steps today", but we ask that you not mention that you were asked to tell others about it (either in person or digitally). *Please also make sure that you have your step tracker turned on for these three days!*
>
> On **Friday**, we will send you a follow-up email to ask the following for each day:
> - Did you walk at least 5,000 steps?
>   - We will ask for a screenshot or other image of your step count as verification.
> - How many people did you tell about the task?
> - What method did you use to complete your communication: digital, in-person, or a mix?
>
> **Once we receive your final data, you will be entered into a raffle for one of twenty $25 Amazon gift cards** (regardless of whether that data says you walked as much as

we asked you to or not!). If you do not share your completion data with us, then you will not be eligible to receive a gift card.

If you have any questions or concerns, please feel free to reach out to us! Otherwise, we will follow up with you on Friday to see how this week went and wrap up the study.

Thanks so much!

Jenni (jennifer.m.podracky@berkeley.edu)
April (mjapkim@berkeley.edu)
Saurav (sauravdatta@berkeley.edu)

Please note that this is the messaging shared for the full study participants; pilot study recipients received a similar communication.

Data Collection Survey

Upon study completion, subjects were sent a survey designed to check their compliance with their assigned treatments, and to collect their outcome measurements. The data collection survey was structured as follows:

Question Purpose	***Question Text***
Informational	**Congratulations on completing the study!** You were asked to complete a task each of the last three days (Tuesday, Wednesday, and Thursday), and were also potentially asked to communicate with others about this task. This form is to collect your results; for each day of the study (Tues-Thurs), you will be asked: ● How many people you told (if any) ● What communication method(s) you used (if any) ● How many steps you took ***Please be honest - it's much more important that you be truthful than whether you completed your assignment or not!*** **Upon submission of your results, you will be entered into a raffle for one of twenty $25 Amazon gift cards** (regardless of whether that data says you walked as much as we asked you to or not!). **We will stop accepting submissions at 11:59 PM on Monday, December 10th**, so please submit your results before then to be considered! If you are a gift-card winner, we will contact you by Friday of next week (December 14) to notify you. Otherwise, we thank you for participating in our study!

	Jenni (jennifer.m.podracky@berkeley.edu) April (mjapkim@berkeley.edu) Saurav (sauravdatta@berkeley.edu)
Informational	**Please be assured that the details you share will not be released or published in any form outside of this study.**
Contact	What is your name?
Contact	What is the email with which you signed up for the study?
Covariate	Do you personally know one of the individuals conducting the study?
Compliance Check	This is for the activity that you were assigned to for Tuesday. How many others did you inform, if at all, of your intent to complete the task? ● None ● 1 ● 2 or more
Compliance Check	This is for the activity that you were assigned to for Tuesday. How did you communicate, if at all, to your friends/family members of your intent to complete the task? ● Not applicable ● In person ● Through digital means ● Both in person and through digital means
Outcome	How many steps did you take on Tuesday?
Outcome	Please upload a screenshot showing the steps taken for Tuesday. You don't have to upload anything if you did not do the task or were not instructed to do a task.
Compliance Check	This is for the activity that you were assigned to for Wednesday. How many others did you inform, if at all, of your intent to complete the task? ● None ● 1 ● 2 or more
Compliance Check	This is for the activity that you were assigned to for Wednesday. How did you communicate, if at all, to your friends/family members of your intent to complete the task? ● Not applicable ● In person ● Through digital means ● Both in person and through digital means

Outcome	How many steps did you take on Wednesday?
Outcome	Please upload a screenshot showing the steps taken for Wednesday. You don't have to upload anything if you did not do the task or were not instructed to do a task.
Compliance Check	This is for the activity that you were assigned to for Thursday. How many others did you inform, if at all, of your intent to complete the task? ● None ● 1 ● 2 or more
Compliance Check	This is for the activity that you were assigned to for Thursday. How did you communicate, if at all, to your friends/family members of your intent to complete the task? ● Not applicable ● In person ● Through digital means ● Both in person and through digital means
Outcome	How many steps did you take on Thursday?
Outcome	Please upload a screenshot showing the steps taken for Thursday. You don't have to upload anything if you did not do the task or were not instructed to do a task.

Please note that questions with bullets listed beneath them are multiple choice; the bullets list the possible answers that participants had to choose from. All questions other than the multiple choice questions accepted a freeform text response.

Covariate Encodings

Gender	
0	Male
1	Female
2	Gender non-conforming
3	NA
Age Range	
0	18 - 24

1	25 - 34
2	35 - 44
3	45 - 54
4	55 - 64
5	65+
6	NA
Living Situation	
0	Lives alone
1	Lives with others
2	NA
Knows an Author	
0	Does not personally know any of the authors
1	Personally knows at least one of the authors
2	NA

Table 9. Covariate encodings

Regressions

	Dependent variable:
	Steps > 5000
Social commitment	−0.047 (0.064)
Constant	0.333 (0.260)
userId3	0.667** (0.266)
userId6	1.000*** (0.018)
userId13	1.000*** (0.018)
userId14	0.333 (0.260)
userId17	0.333 (0.260)
userId19	1.000*** (0.018)
userId22	0.667** (0.266)
userId25	1.000*** (0.018)
userId26	1.000*** (0.018)
userId28	1.000*** (0.018)
userId33	1.000*** (0.018)
userId39	1.000*** (0.018)
userId45	1.000*** (0.018)
userId47	0.333 (0.279)
userId54	1.000*** (0.018)
userId56	1.000*** (0.018)
userId57	1.000*** (0.018)
userId58	0.667** (0.285)
userId59	1.000*** (0.018)
userId65	1.016*** (0.025)
userId66	0.667** (0.266)
userId68	1.000*** (0.018)
userId69	0.667** (0.285)
userId73	0.667** (0.285)
userId75	1.000*** (0.018)
userId77	1.000*** (0.018)
userId84	1.000*** (0.018)
userId85	1.000*** (0.018)
userId86	0.667** (0.266)
userId88	1.000*** (0.018)
userId91	1.000*** (0.018)
Constant	0.031 (0.045)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 10. Full primary linear regression coefficient estimates using two levels of treatment

	Dependent variable:
	Steps > 5000
Digital commitment	−0.001 (0.065)
In person commitment	−0.092 (0.076)
Constant	0.333 (0.261)
userId3	0.667*** (0.248)
userId6	1.000*** (0.035)
userId13	1.000*** (0.035)
userId14	0.333 (0.261)
userId17	0.333 (0.261)
userId19	1.000*** (0.035)
userId22	0.667*** (0.248)
userId25	1.000*** (0.035)
userId26	1.000*** (0.035)
userId28	1.000*** (0.035)
userId33	1.000*** (0.035)
userId39	1.000*** (0.035)
userId45	1.000*** (0.035)
userId47	0.333 (0.298)
userId54	1.000*** (0.035)
userId56	1.000*** (0.035)
userId57	1.000*** (0.035)
userId58	0.667** (0.287)
userId59	1.000*** (0.035)
userId65	1.016*** (0.046)
userId66	0.667*** (0.248)
userId68	1.000*** (0.035)
userId69	0.667** (0.287)
userId73	0.667** (0.287)
userId75	1.000*** (0.035)
userId77	1.000*** (0.035)
userId84	1.000*** (0.035)
userId85	1.000*** (0.035)
userId86	0.667*** (0.248)
userId88	1.000*** (0.035)
userId91	1.000*** (0.035)
Constant	0.031 (0.049)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 11. Full primary linear regression coefficient estimates using three levels of treatment