Complex Contagions in Charitable Donations

Jie Gao

Dept. of Computer Science, Stony Brook University, Stony Brook, NY

Golnaz Ghasemiesfeh

Dept. of Computer Science, Stony Brook University, Stony Brook, NY

Jason J. Jones

Dept. of Sociology and Institute for Advanced Computational Science, Stony Brook
University, Stony Brook, NY

Grant Schoenebeck

School of Information, University of Michigan, Ann Arbor, MI

Author Note

The authors have chosen to be listed in alphabetical order. Corresponding author:
Jason J. Jones Jason.J.Jones@stonybrook.edu

Abstract

Cascades over social networks can spread information, beliefs, diseases, technologies, and behaviors. Simple cascades spread from mere contact and produce submodular influence curves. Complex cascades assume agents with thresholding behavior and may produce non-submodular influence curves. In this study, we run three experiments that request charitable donations from human participants and experimentally manipulate whether and where their peers donate. We find evidence that we can (1) direct donations to an otherwise unpopular charity and (2) elicit complex contagion as evidenced by a non-submodular influence curve. The findings represent the most straightforward evidence to date of treatment-induced complex contagion - explicitly and formally defined - in human decision-making.

Keywords: social influence, complex contagion, charity, donations, experiment

Complex Contagions in Charitable Donations

Introduction

Human activity is embedded in a network of social interactions, which can spread information, beliefs, diseases, technologies, and behaviors. A better understanding of these social interactions promises a better understanding of and the ability to influence a wide range of phenomena, e.g., financial practices (Banerjee, Chandrasekhar, Duflo, & Jackson, 2013; Coleman, Katz, & Menzel, 1957), healthy/unhealthy behaviors (Centola, 2018; Mermelstein, Cohen, Lichtenstein, Baer, & Kamarck, 1986), and voting habits (Bond et al., 2012; Jones, Bond, Bakshy, Eckles, & Fowler, 2017). In this work, we examine the effect of social influence on decisions regarding charitable donations.

Many cascade models have been proposed (Arthur, 1989; Morris, 2000; Watts, 2002), and two simple examples are the Independent Cascade model (Kempe, Kleinberg, & Tardos, 2003, 2005; Mossel & Roch, 2010) and the Threshold model (Granovetter, 1978). In each of these models, there is a set of nodes on a network, and some set of these nodes start off “infected”.¹ In the Independent Cascade model, each newly infected node infects each currently uninfected neighbor in the subsequent round with some fixed probability p . In the Threshold model each node has a threshold (0, 1, 2, etc.) and becomes infected when the number of infected neighbors meets or surpasses that threshold.

Cascades that spread according to different models may behave radically differently. Thus, to understand the way cascades spread requires using the correct models for these cascades. An especially important distinction, is whether cascades are simple (e.g. independent cascade model) or complex (e.g. threshold model), a definition we will formalize soon. Unfortunately, teasing out the forms of cascades requires isolating the impact of influence from other confounding factors such as homophily and environmental effects, which is incredibly difficult to do with observational data. This

¹ We keep the usual notation of infected versus uninfected. “Adopters” versus “non-adopters” is perhaps more precise for most of the settings we are concerned with, but is considerably more clunky.

motivates our experimental studies.

Complex Cascades

Centola and Macy in their seminal work “Complex Contagions and the Weakness of Long Ties” (Centola & Macy, 2007) proposed to better understand the myriad models of cascades by dividing them into just two categories: simple and complex. The differentiating feature of complex contagion, compared to simple contagions is that complex contagions require reinforcement from multiple social contacts, while simple contagions only require one contact.

The definition of complex contagions recognizes an important distinction between the acquisition of information and the decision to act on the information. While it takes only a single tie for people to hear about a new belief, technology, fad or fashion, “it is when they see people they know getting involved that they become most susceptible to recruitment” (Centola & Macy, 2007). Thus, while the former can be suitably modeled by a simple contagion, the latter is better modeled by a complex contagion, in which multiple sources of activation are required because contact with a single active individual is not enough to trigger adoption. Suggested features of decision contexts that lead the contagion process to be complex include strategic complementarity, emotional exchange, and uncertainties surrounding credibility and legitimacy (Centola & Macy, 2007) and thus apply especially to behaviors involving great risk (e.g. adopting pricey technologies, emigration, etc.).

Mathematically, a cascade from the general threshold model can be represented as a tuple $G, \mathbf{f}, S$ where G is an undirected graph with vertices V and edges E ; $\mathbf{f} = \{f_v\}_{v \in V}$ is a collection of influence functions $f_v : \{0, 1\}^{\Gamma(v)} \rightarrow [0, 1]$,² and S is a set of initially infected vertices. Each node has a label $\ell(v) \in \{0, 1\}$, and a threshold θ_v from $[0, 1]$. Each node $v \in S$ starts with $\ell(v) = 1$ and is called “infected”, and the remaining nodes begin labelled $\ell(v) = 0$ and are called “non-infected”. A non-infected v node becomes infected if the influence of its infected neighbors is greater than its threshold: that is

² We use $\Gamma(v)$ to denote the neighbors of v in the network.

$f_v(\ell(\Gamma(v))) > \theta_v$. We denote the expected number of infected vertices with seed set S as $\sigma(S)$. Note that $\sigma : 2^N \rightarrow \mathbf{R}$.

A function $f : 2^N \rightarrow \mathbf{R}$ is submodular if $f(S \cup T) + f(S \cap T) \leq f(S) + f(T)$ for all $S, T \subseteq N$. Intuitively this definition states that the marginal contribution of each element decreases as more elements are added. If f depends on the the *number* of elements in its input, then this is to say that $f(k)$ is concave.

We say that a cascade is *simple* if the expected number of infections $\sigma(\cdot)$ is submodular, and we say that a cascade is *complex* if the expected number of infections $\sigma(\cdot)$ is non-submodular. Intuitively, a contagion is *simple* when the influence of an additionally infected neighbor is marginally decreasing and is *complex* when this fails to hold (e.g. contagions that require activation from multiple neighbors). In simple contagions, there is no synergy. A node’s influence is only eroded by other nodes’. The Independent Cascade model is a simple cascade because a vertex’s marginal probability of becoming infected after a new neighbor is infected decreases with the number of previously infected neighbors (Kempe et al., 2003). In contrast, the Threshold model (with $k > 1$) is a complex cascade because the marginal probability of being infected may increase as more neighbors are infected. For example, if a vertex has a threshold of 2, then the first infected neighbor has zero marginal impact, but the second infected neighbor causes this vertex to become infected with probability 1.

Here f is based on the number of infected neighbors. We will necessarily average over all the nodes. We note that if all the $f_v(\cdot)$ are sub-modular, then their average is as well. This is because the set of submodular functions is convex. Conversely, if the “average” of the influence functions f is non-submodular, then some of the f_v ’s must be non-submodular. Figure 1 graphically represents submodular and non-submodular influence curves.

Evidence for the Existence of Complex Cascades

There have been few randomized experiments conducted to study complex contagion. A notable exception is Centola (2010) in which subjects were recruited and

randomly assigned to two groups. Each group was placed upon a pre-specified network topology with one network being a random network and the other one being a clustered lattice network. Diffusion was initiated by a random seed sending signals to its network neighbors encouraging them to adopt a health-related behavior (registering for a health forum website), and each adopting neighbor triggered messages sent to her neighbors. Complex contagions spread wider and faster in the clustered lattice networks and spread slowly, or stopped spreading, in the random networks. The difference across the two networks in the number of nodes that eventually adopted provided indirect evidence that the contagion process was complex.

A “natural experiment” was examined by Aral and Nicolaides (2017). They examined the contagion of running among friends using weather as an exogenous variable. However, they did not find that complex contagions were a good fit for their data.

There have been many *observational* studies measuring contagions over networks. Scores of these are surveyed in a recent work by Guilbeault, Becker, and Centola (2017). Many examples of complex contagions have been reported, including the purchase of expensive technology, change in social behavior, and the decision to participate in a migration (Coleman et al., 1957; Macdonald & Macdonald, 1964). In Christakis and Fowler’s work (Christakis & Fowler, 2008), evidence was found that the likelihood a smoker will quit depends on their exposure to multiple contacts. This was confirmed by a more recent study by Kuhlman et al. (2011), and on QuitNet (Myneni, Cobb, & Cohen, 2013). Another example of complex contagions is Davis’s (Davis & Greve, 1997) analysis of the spread of “poison pills” through the network of cooperate board membership in response to the rise in cooperate takeovers during the 1980s.

In computer science, studies of large scale data sets from online social networks observe traces of complex contagions. For example, studies on Facebook discovered that having multiple friends already on Facebook substantially increased the likelihood of one joining Facebook (Ugander, Backstrom, Marlow, & Kleinberg, 2012). A study on the Twitter platform shows a similar phenomenon, that *persistence* (the effect of

repeated exposure to a topic) plays an important role in the diffusion of hashtags (Romero, Meeder, & Kleinberg, 2011).

Unfortunately, there are severe limitations to what can be deduced from observational studies. In such studies, a cascade is observed and then the (empirical) likelihood that an agent with k infected neighbors becomes infected is reported. The biggest problem is a form of selection bias: the nodes that have k -neighbors infected are not a random set of nodes. It could be that (due to homophily³) nodes that are a priori more likely to adopt, are also more likely to have many adopting neighbors. An additional problem is that such studies typically assume that their view of the network is complete and cannot account for information transmission that is likely to be occurring along unobserved edges.

One way to overcome the aforementioned challenges is to run controlled experiments, in which users are recruited and partitioned randomly into groups with different treatments in which their neighbors' attributes and behavior are specified by the experimenters. Such experiments eliminate the selection bias of ties that form in natural networks, allow the experimenter to control the network, and reduce other confounding factors due to random assignment.

Literature on Charity Donation

Behavioral cascades operate in many decision-making contexts. In this work, we chose to study charitable donation decisions with unfamiliar but deserving organizations as targets, because we desired a task in which subjects made decisions about their own money in a context of social information and no clear correct answer.

There has been much work looking into the mechanisms of charitable giving decisions (Bekkers & Wiepking, 2011). Some characteristics affecting donation behavior have been studied by data collection through surveys (Finkelstien, 2009; Lee & Chang, 2007). For example, it was shown that determinants affecting volunteering for charities

³ Homophily is a commonly observed social phenomena that like-minded persons tend to be connected more often than two random persons.

(donating time) are mostly intrinsic attributes while those for monetary donations are mostly extrinsic (Lee & Chang, 2007). Personal attributes and characteristics were also shown to correlate with the choice of charities (Bennett, 2003). Additional factors such as the price, donation matching, or seed fund were studied as well (Gneezy, Gneezy, Nelson, & Brown, 2010; Karlan & List, 2007; List & Lucking-Reiley, 2002).

In DellaVigna, List, and Malmendier (2012) the authors studied two possible motivations to charitable giving: altruism and social pressure (people would rather not give but dislike saying no). They demonstrated that at least some giving is driven by social pressure. When a door-to-door fundraiser was altered so it was easy to avoid the donation request, the number of contacts and total donations decreased.

The contagious behavior of charity donations was studied in Korolov et al. (2015) in which the level of ‘chatter’ on Twitter and the amount of donation were compared. The hypothesis was that people who donated may also tweet about it, and the tweets may trigger other people to donate (who may remain silent), indicating a possibly super-linear relationship of the chatter and the amount of donation received. By using observational data, it was estimated that the amount of donation grows as a quadratic function of activity levels on Twitter. However, the direction of causality cannot be determined with correlational results. In a different study (Lacetera, Macis, & Mele, 2016), it was shown that broadcasting a pledge on social media is associated with more pledges by a user’s contacts, but the subsequent boost in donation amount was minimal. Although there was considerable activity in the forms of clicks and “likes,” only a small number of donations (30 out of 6.4 million users reached) were made.

Our work contributes to the literature by demonstrating the contagious properties of charity donation in a controlled experimental setting.

Hypotheses and Contributions

In this paper we ran experiments to test the following three hypotheses.

- Hypothesis I: Subjects will be influenced by their peers. That is, the more peers give, the more likely the subjects will give.

- Hypothesis Ia: Peers who are perceived as similar will be more influential than peers who are perceived as dissimilar.
- Hypothesis II: Contagions will be complex. That is, the influence curves will be non-submodular.

To test these hypotheses, we designed and ran three experimental studies. In each experiment, each participant was rewarded monetarily for answering 10 questions. They were then asked if they wanted to donate a bonus payment to one of a given list of charities. They were informed that they were put in a ‘group’ with others simultaneously participating, and they observed what their groupmates (who were actually bots that we controlled) did. Using this methodology, we could explore how influenced subjects were by their network neighbors’ actions. In particular, we manipulated the donation behaviors of the neighbors – specifically, we varied how many neighbors donated to a charity, how many kept the money and how many chose a particular charity that was otherwise disfavored.

We ran three experiments of minor variations in design, with a total of 4474 participants from Amazon Mechanical Turk. The main observations from the experiments were the following. When more neighbors decide to donate (to any charity), a subject is more likely to donate. Thus charity donation is indeed contagious. This confirms Hypothesis I. In addition, when multiple neighbors decided to donate to the *same* charity, the chance that a subject donated to that particular charity was significantly lifted. Somewhat surprisingly, we did not find evidence that the influence from neighbors who agreed with the participant in attributes/opinions was much different from the influence from disagreeing neighbors. Thus we do not have evidence to support Hypothesis Ia.

For Hypothesis II, we did not find evidence of complex contagion in whether or not participants give. Yet, we do find consistent support for Hypothesis II in whether participants give *to a particular charity*. We argue that complex contagion emerges when the space of options is large causing the chance of any particular option being adopted to be small, and social signals thus take on added salience. These findings are

of theoretical importance to understanding complex contagion and of practical importance to organizations wishing to maximize donations.

Experiment A

Experiment A Methods

In this section, we present a full account of the protocol for the first experiment. For subsequent experiments, we describe only the difference from the previous protocol. The experiment protocol was approved by the Stony Brook University Institutional Review Board.

The experiment was conducted online through Amazon Mechanical Turk. Subjects interacted with a web application that presented the stimuli, recorded responses and simulated behavior of other users. The experiment had four phases:

1. Personal preference questions;
2. Group formation;
3. Donation choice;
4. Survey and demographic questions.

The focal independent variables (Neighbor Similarity and Neighbor Donation Choices) were delivered in Phase 2 and 3 and the dependent variable (Donation Choice) measured in Phase 3.

Phase 1: Personal Preference Questions. The experiment began with ten questions regarding personal preference. Each question had two possible answers, and the subject had to choose one or the other. The experiment would not move on until an option was selected in response to each question.

Our goal in this phase of the study was to set up the Similarity manipulation that would occur in the next phase. Specifically, we aimed to make the subject feel that they were similar to or different from their cohort of neighbors by first asking them about their preferences and then manipulating to what degree their neighbors shared those

preferences. Because the “neighbors” were simulated, we could randomly assign preference similarity.

We used 10 questions culled from online lists of good questions to ask when first getting to know someone. For example, “Do you prefer books or movies?”, “Do you prefer individual sports or team sports?” The full list of questions is included in the appendix.

Phase 2: Group Formation. In this phase, the subject was told they were being connected to other subjects to form a group. All subjects were told

- They would make a decision about donating a bonus payment in the next phase.
- They would be assigned to a group with three other participants.
- Their groupmates might or might not have answered the 10 questions similarly.
- The donation choices of their groupmates would be visible, but their own choice would be private.

The Similarity manipulation was delivered at this point. The application populated the groupmate slots with neighbors from one of two categories. Either all three neighbors were labelled Low Similarity or all were labelled High Similarity. Low Similarity was defined as “Less than 3 answers in common with you.” High Similarity was defined as “More than 7 answers in common with you”. The neighbors were displayed as lettered avatars (A,B,C). Figure 2 contains an example screenshot depicting Phase 2.

Phase 3: Donation Choice. In this phase, the Neighbor Donation Choices treatment was delivered, and the subject made their donation decision. Specifically, the subject was asked “Where would you like to donate?” and was presented with the names of many charities. The charities were real and were selected by browsing online databases of highly-rated charities. The subjects were promised \$0.50 for participating in the experiment and always paid at least this amount. Within the experiment they were then told they had earned \$0.60 and could choose to donate or keep the “bonus” \$0.10.

In the experiment, a list of 15 charities was displayed in a matrix with an additional option to “Keep 10 cents and Do Not Donate.” The participants putative groupmates’ decisions were revealed as the participant considered the options. To make the neighbors appear more like real persons to the subject, the donation choices of the neighbors were delayed in display with a random time duration. At approximately 5, 7.5 and 9 seconds after the screen loaded, a simulated neighbor made a charity decision. (The exact times were randomly jittered up or down up to 1 second.)

The condition the subject was randomly assigned determined the neighbors’ donation choices. Of most interest was the number of neighbors who chose to donate to the charity named Schistosomiasis Control Initiative (SCI). 0, 1, 2 or all 3 neighbors chose SCI, thus allowing a full characterization of the influence curve from no seeded infection to complete infection.

The subject was required to make a choice before continuing. This choice is the focal dependent variable. Figure 3 contains an example screenshot depicting Phase 3.

Phase 4: Survey and Demographic Questions. This phase included a short set of demographic and survey questions which we felt might be useful in post-hoc exploration of the experimental influence results. The questions followed the delivery of the treatments and the collection of the dependent variable so as to not create any demand characteristics within the experiment. Figure 4 contains an example screenshot depicting Phase 4.

Experiment A Results

Basic Statistics. We recruited 1435 participants from Mechanical Turk and assigned them randomly to 14 conditions (Table A1 in the Appendix) which specified whether their neighbors were similar or dissimilar and what charity choices their neighbors made. All three neighbors had the same similarity - either all High or all Low. In addition to the Similarity variable, the conditions differed by the choices made by the three neighbors: keeping the money to self (Self); donating to Schistosomiasis Control Initiative (Schis), or donating to charities other than SCI chosen at random

without replacement (Rand). We continued recruiting subjects and assigning conditions randomly until every condition contained at least 100 complete submissions.

We recorded whether the participants chose to keep the money and if the participants decided to donate, which charity was chosen. Summary results are shown in Table 1. As can be seen from the data, a majority of the participants (885 out of 1435, 61.67%) decided not to donate and kept the 10 cents. Generally speaking, charities with easily-understood names received more donations. To illustrate, the top 3 charities are respectively Project Healthy Children, Centre for HIV and AIDS Prevention Studies, and Global Fund to Fight AIDS. The bottom 3 charities are END Fund, Deworm the World Initiative led by Evidence Action, and Measles and Rubella Initiative. The Schistosomiasis Control Initiative, the charity we chose to highlight with our experimental manipulation, received a donation from 30 participants - about 2.1% of participants.

Testing for Social Influence (Hypothesis I)

We use our experiment to test Hypothesis I in two different ways. First, we measured whether subjects who were exposed to neighbors that donated their 10 cents bonus would also donate. Second, we measured whether subjects who were exposed to neighbors that gave to the Schistosomiasis Control Initiative (SCI) would also donate to SCI. SCI was chosen as the target charity for the influence manipulation because it was among the least popular in a pilot test.

Influence for Donating versus Keeping. We first look at the fraction of participants who chose to donate to any charity (i.e. chose any option other than Keep 10 cents and Do Not Donate), when they have 0, 1, 2, 3 neighbors who have chosen to donate. Figure 5 (Top) shows the results. Clearly, the more neighbors donating, the more likely a subject will donate.

Logistic regression. We ran logistic regression on the data. The dependent variable is binary, with 1 denoting that the subject donated to a charity and 0 denoting the subject chose Keep 10 cents and Do Not Donate. We define the independent

variables as the number of neighbors donating to a charity, the similarity of the neighbors and an interaction. The coefficients and p -values of the logistic regression are shown in Table 2. For the purpose of the present discussion, we point out that having neighbors who donated to some charity clearly shows influence, with a p -value of 0.0275.

Influence for Donations to Schistosomiasis Control Initiative. Here we look at the percentage of participants who chose to donate to SCI when they have 0, 1, 2, 3 neighbors who have donated to SCI. Figure 5 (Bottom) shows the results. As can be seen from the data, the more neighbors who donate to SCI, the more likely a participant will also donate to SCI. With no neighbors donating to SCI, only .48% of the subjects chose to donate to SCI. With one, two, or three neighbors donating to SCI, 1.1%, 2.7%, and 5.4% of the subjects chose to donate to SCI. Thus, with three neighbors the observed donation rate was over eleven times greater than the baseline rate of choosing SCI in the absence of peers choosing it.

Again, we ran a logistic regression on the data (Table 3). The target variable is binary, with 1 meaning that the subject donated to SCI and 0 otherwise. The logistic regression points to a statistically significant effect for the number of neighbors choosing SCI (p -value of 0.0053).

Hypothesis Ia. To test Hypothesis Ia we separated participants by those who were presented with High Similarity (HS) neighbors and those with Low Similarity (LS) neighbors. See Figure 6 (Left) for the influence curves for SCI.

The logistic regression included the Similarity treatment as an independent variable. We focus on the interaction between neighbor similarity and neighbor influence to evaluate Hypothesis Ia. If there was a negative interaction, this would mean that LS neighbors have a weaker effect than HS neighbors. The interaction terms are included in the regression both for donations in general (Table 2), and donations to SCI in particular (Table 3).

Somewhat surprisingly, there is no discernible interaction effect. High and Low Similarity neighbors have influence of similar direction and magnitude. Because we do not find a reliable effect of this manipulation, we collapse across the Similarity variable

(HS and LS conditions) in our analyses.

Testing for Complex Contagion (Hypothesis II)

To test Hypothesis II, we seek evidence that the influence curve is non-submodular. We evaluate this in two contexts: donating at all, and donating to a specific charity.

Looking at Figure 5 (Top), there is no visual evidence of non-submodularity. It appears the second or third donating neighbor is not substantially more effective than the first in persuading a subject to donate. In contrast, in Figure 5 (Bottom) the visual evidence suggests each additional neighbor exerted stronger influence than the previous, indicating that the influence from having neighbors donating to SCI is non-submodular. We fit the curve by a linear function and an exponential function to show the contrast.

Testing Significance. It is not a priori clear how to test if a function is submodular. However, we might start by observing that a submodular function should satisfy a suite of tests. Thus, a non-submodular function is one that violates one of these tests. Here we formalize constraints that submodular functions shall satisfy.

Let $P(i)$ be the probability of donating conditioned on being in a treatment where i neighbors donate. If the influence curve is submodular, 5 different constraints should be satisfied:

$$P(2) - P(1) \leq P(1) - P(0)$$

$$P(3) - P(2) \leq P(2) - P(1)$$

$$P(3) - P(2) \leq P(1) - P(0)$$

$$P(3) - P(1) \leq 2(P(1) - P(0))$$

$$2(P(3) - P(2)) \leq P(2) - P(0)$$

Note that the last two constraints are implied by the first three. The difficulty in translating this definition into a statistical test is correcting alpha for multiple comparisons. Moreover, the tests share inputs and are likely to be correlated.

Fortunately, Chen and Szroeter (2009) provides a method to handle this situation. Their procedure takes in multiple tests, and outputs the p -value of the event

that at least one of the tests is violated. To be precise, the null hypothesis is that none of the tests are violated. The p -value then represents a bound on the probability of the data, given that none of the tests are violated. The procedure explicitly takes into account the covariance of the tests, and the paper includes a proof the test is asymptotically exact. We provide details regarding our application of the procedure and our calibrations with simulated data in the appendix.

For donating at all, the Chen-Szroeter test produces a p -value of 0.4667. For donating to SCI, the Chen-Szroeter test produces a p -value of 0.186. We interpret this as showing some support of non-submodularity in the case of SCI, though it is far from conclusive.

Experiment A Discussion

In Experiment A, we found evidence that subjects respond to social influence. Specifically, they are more likely to donate their bonus if they observe others do so, and are more likely to choose a specific charity the more of their peers they observe choosing that charity. In the case of donating to SCI, the shape of the influence curve appears non-submodular - in other words, there are increasing returns to additional peers choosing that charity, but a formal test provides only weak support this is a reliable effect.

In the next experiment, we aimed to replicate these results. To probe the lack of effect of the Similarity manipulation we made one change to increase its salience.

Experiment B

Experiment B Methods

Experiment B follows nearly the same design as Experiment A. However, we selected new questions for Phase 1: Personal Preference Questions in order to increase the salience of these items and move from a subtle manipulation of similarity to one more intense. In Experiment A, Phase 1 contained ten getting-to-know-each-other ‘vanilla’ questions. In Experiment B, we substituted ten questions drawn from surveys

revealing politically polarized responses. For example, “Do you believe the United States would be better if it were a more Christian nation?” The full list of questions is included in the appendix.

We recruited 1443 participants from Mechanical Turk for this experiment, partitioned in the same 14 conditions, with more than 100 participants per condition, similar to Experiment A.

Experiment B Results

Summary results are shown in Table 4. Again, keeping the bonus and charities with easily understood names were popular choices. The rank popularity of options closely matches that of Experiment A.

Table 5 presents coefficient estimates for a logistic regression with donating to any charity as the target variable. Just as in the previous results, subjects are more likely to donate to any charity the more neighbors they see donating. There is no evidence the strength of this effect is mediated by similarity.

Table 6 presents coefficient estimates for a logistic regression with donating to SCI as the target variable. Subjects are more likely to donate to SCI the more neighbors they see donating to SCI. There is no evidence the effect is mediated by similarity.

Figure 7 depicts influence curves for subject donation probability as a function of donating neighbor count. The curve for donation to SCI is clearly non-submodular. Visually, we show an exponential curve fits the data more closely than a line. Again we ran the Chen-Szroeter test to measure the statistical significance of this result. The test results in a p -value of 0.017. This shows strong evidence of non-submodularity.

Experiment B Discussion

The results of Experiment B were very similar to those for Experiment A. Hypothesis I is strongly supported - subjects are more likely to donate and donate specifically to SCI as the number of peers they are observing taking these actions increases. Again, the data show no support for Hypothesis Ia. Despite our effort to increase the salience and pertinence of the similarity manipulation by using politically

divisive questions, the effect of similarity and its possible interaction with the slope of the influence curve are statistically indistinguishable from zero. Hypothesis II again receives its strongest support in the context of donations to SCI. Subjects who observe 3 neighbors donate to SCI are 11 times more likely than subjects who observe none to also donate to SCI, and the influence curve is non-submodular.

Experiment C

Experiment C Methods

In Experiment C we made additional changes to the experiment conditions.

1. *Increased neighbors to 5.* In Experiment A and B, only three neighbors were presented to the participants. When three (out of three) neighbors donate to the same charity, there are two manners in which this differs from two (out of three): 1) one additional neighbor donated and 2) now all neighbors have unanimously made the same choice. Similarly, when two neighbors are observed making the same choice, there is a conflation of two possible effects - a count of two and a majority of the group. To separate the effects with minimal change, in Experiment C we increased the count of neighbors to five, while only up to three donate to the target charity. This is the key distinguishing feature of Experiment C.
2. *Everyone donates.* In Experiment A and B participants have the option of keeping the 10 cents. In Experiment C, we removed the option of keeping the money. Our goal was to increase statistical power.
3. *Quiz on Group Phase Instructions.* The participants must answer five True/False questions regarding the Group Phase Instructions correctly before they were allowed to move on. This was to make sure participants completely understand the instructions.
4. *New bot response timing.* Due to the large number of neighbors in Experiment C, the response time was chosen randomly from a uniform distribution over the range [8000, 20000] milliseconds. This range covers over 80% of the empirically observed

response times from Experiments A and B. We use this larger range in response times to avoid the neighbor choices clustering in a short, early interval. Without this change, the neighbors responses appeared artificial to subjects, according to feedback on an unused pilot set of subjects. Toward the same goal of presenting a realistic experience of simultaneous users, we also informed the subjects that their groupmates were deliberately chosen so as to all have high agreement or all have high disagreement with them.

5. *Emphasizing disagreement.* In Experiment C, we continued with ten politically polarized questions. We changed the wording subjects saw from “High Similarity” to “High Agreement” and “Low Similarity” to “High Disagreement”. We again hoped to increase the salience of the Similarity manipulation.

We ran Experiment C with 1596 participants from Mechanical Turk, randomly assigned into 8 conditions: Similarity (High or Low) x SCI Neighbors (0, 1, 2, 3).

Experiment C Results

Because all participants in Experiment C donate, we only examine how many of them donate to SCI. The results can be seen in Figure 8. When zero neighbors chose SCI, only 0.24% of the participants donated to SCI, yet with 3 out of 5 neighbors donating to SCI, the rate is elevated to 5.79%, which is 24 times greater. We fit the influence curve by both a line (in blue) and an exponential (in red). The Chen-Szroeter test returns a p -value of 0.21, indicating that violations of submodularity are more likely than not.

Table 7 presents coefficient estimates for a logistic regression with donating to SCI as the target variable. Subjects are more likely to donate to SCI the more neighbors they see donating to SCI. There is no evidence the effect is mediated by similarity.

Experiment C Discussion

Again the data support Hypotheses I and II but not Ia. Participants are more likely to donate to SCI the more neighbors they observe doing so, regardless of the

reported agreement in their answers to politically polarizing questions. The average influence function would likely be non-submodular.

Exploratory Analysis - Who is more likely to donate?

Post-hoc, we explored the other recorded variables as predictors in larger models predicting which subjects would donate rather than keep their bonus.

We find that:

- People who participated for fun are more likely to donate than people who came for money, with a very small p -value (less than 0.00001).
- Females are more likely to donate than males, with a p -value of 0.0034.
- For the politically polarized questions in Experiment B and C, we observe that certain questions show statistical significance on whether the subject will donate. We also scored each the answer as liberal or conservative and summed the values to form a scale from 0-10. The scale did not show a significant effect as a predictor.

General Discussion

These results comport with and extend previous findings of social contagion. We find that our subjects conform with the reported behaviors of others when choosing whether or not to donate their earnings. We find that peer choices can powerfully direct subjects to a particular charity (SCI) which is an otherwise unpopular option. Most strikingly, we demonstrate complex contagion in a true randomized, controlled experiment as evidenced by the non-submodularity of the influence curve for SCI adoption in all three experiments (Figure 5, 7, 8). We calibrated and ran a test for non-submodularity by adopting the methods of Chen and Szroeter. While the significance of the non-submodularity was modest in Experiments A and C, Experiment B found strong statistical evidence for non-submodularity.

These results should encourage further theoretical and experimental work exploring complex contagion. We believe that characterizing influence curves as

submodular or not is a useful distinction. Complex contagions will reveal themselves as non-submodular functions. We find some evidence for such when human subjects are making otherwise unpopular (low base rate) decisions such as donating to a charity with an obscure name, but not when subjects are making the popular decision to keep their money for themselves. This comports with previous theory - that complex contagions should be observed for adoption of risky or unproven options, but further research could more directly manipulate the perceived risk or baseline popularity of available options.

In pilot testing, we found that binary decision formats (i.e. donate to SCI or do not) were likely to result in linear influence curves. When we expanded the option space, we observed the non-submodularity demonstrated in these results. This is an important theoretical consideration. Most work on social contagion frames contagions as binary: Did the doctor prescribe the new drug or not? Did the citizen vote or not? The power of social information, however, may lie in the multiplicity of alternatives inherent in real-life choices. Complex contagion may be better understood if behaviors are re-contextualized from X or Not-X to choosing X from the broad array of all X and non-X possibilities. A limitation of our results is that we cannot distinguish the effect of number of options versus popularity of the target option prior to influence. Further research which manipulates social signals in the context of decisions with varying numbers of options would be valuable.

Practical guidance for maximizing donations to a particular charity also emerges from these results. If naming or re-naming a charity, one should pick a name that is easy to understand and has positive connotations. The most popular charity choices among our subjects follow this rule - e.g. Project Healthy Children - and the least popular do not - e.g. END Fund. Regardless of a charity's name, "social proof" Cialdini (2007) is a strategy that will increase propensity to donate. With the exception of gender and the subject's stated desire to make money, demographics and personal preferences were weak predictors of donation behavior. In contrast, faceless avatars signalling the choices of unknown individuals had powerful and compounding positive effects. The probability

a subject chose to donate to SCI jumped from between $0.2\% - 0.5\%$ to between $5\% - 12.9\%$ with three endorsements - a 10 – 20-fold increase. Soliciting donations with social proof seems more likely to yield benefits than demographic targeting.

Future work should explore these practical considerations and the theoretical importance of baseline popularity and option space size to complex contagion.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant No. CCF-1535912, CCF-1535900, CNS-1618391, and DMS-1737812. The authors would like to thank Avrim Blum for useful insights.

Competing Interests. The authors declare no competing interests.

Data Availability. The datasets and analysis scripts will be provided to the editor and reviewers upon request. Upon publication, the datasets and analysis scripts will be made publicly available through author Jason Jones' Harvard Dataverse: <https://dataverse.harvard.edu/dataverse/jasonjjones>

References

- Aral, S., & Nicolaides, C. (2017). Exercise contagion in a global social network. *Nature communications*, 8, 14753.
- Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*, 99(394), pp. 116–131. Retrieved from <http://www.jstor.org/stable/2234208>
- Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144).
- Bekkers, R., & Wiepking, P. (2011). A literature review of empirical studies of philanthropy eight mechanisms that drive charitable giving. *Nonprofit and Voluntary Sector Quarterly*, 40(5), 924–973.
- Bennett, R. (2003). Factors underlying the inclination to donate to particular types of charity. *International Journal of Nonprofit and Voluntary Sector Marketing*, 8(1), 12–29.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415), 295.
- Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, 329(5996), 1194–1197.
- Centola, D. (2018). *How behavior spreads: The science of complex contagions* (Vol. 3). Princeton University Press.
- Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3), 702–734. Retrieved from <http://www.journals.uchicago.edu/doi/abs/10.1086/521848>
- Chen, L.-Y., & Szroeter, J. (2009). *Hypothesis testing of multiple inequalities: the method of constraint chaining* (Tech. Rep.). cemmap working paper.
- Christakis, N. A., & Fowler, J. H. (2008). The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21), 2249–2258.
- Cialdini, R. B. (2007). *Influence: The psychology of persuasion*. Collins New York.

- Coleman, J., Katz, E., & Menzel, H. (1957). The diffusion of an innovation among physicians. *Sociometry*, 20, 253–270.
- Davis, G. F., & Greve, H. R. (1997). Corporate elite networks and governance changes in the 1980s. *American journal of sociology*, 103(1), 1–37.
- DellaVigna, S., List, J. A., & Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *Quarterly Journal of Economics*, 127(1), 1–56.
- Finkelstien, M. A. (2009). Intrinsic vs. extrinsic motivational orientations and the volunteer process. *Personality and Individual Differences*, 46, 653–658.
- Gneezy, A., Gneezy, U., Nelson, L. D., & Brown, A. (2010). Shared social responsibility: A field experiment in pay-what-you-want pricing and charitable giving. *Science*, 329(5989), 325–327.
- Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6), 1420–1443. Retrieved from <http://www.journals.uchicago.edu/doi/abs/10.1086/226707>
- Guilbeault, D., Becker, J., & Centola, D. (2017). Complex contagions: A decade in review. *arXiv preprint arXiv:1710.07606*.
- Jones, J. J., Bond, R. M., Bakshy, E., Eckles, D., & Fowler, J. H. (2017). Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 us presidential election. *PloS one*, 12(4), e0173851.
- Karlan, D. S., & List, J. A. (2007). Does price matter in charitable giving? evidence from a large-scale natural field experiment. *The American Economic Review*, 97(5), 1774–1793.
- Kempe, D., Kleinberg, J. M., & Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Acm sigkdd* (p. 137-146).
- Kempe, D., Kleinberg, J. M., & Tardos, É. (2005). Influential nodes in a diffusion model for social networks. In *Icalp* (p. 1127-1138).
- Korolov, R., Peabody, J., Lavoie, A., Das, S., Magdon-Ismail, M., & Wallace, W. A. (2015). Actions are louder than words in social media. In *Proceedings of the 2015 ieee/acm international conference on advances in social networks analysis and*

- mining 2015* (pp. 292–297).
- Kuhlman, C., Kumar, V., Marathe, M., Swarup, S., Tuli, G., Ravi, S., & Rosenkrantz, D. J. (2011). A bi-threshold model of complex contagion and its application to the spread of smoking behavior. In *Proceedings of the workshop on social network mining and analysis (sna-kdd 2011)*.
- Lacetera, N., Macis, M., & Mele, A. (2016). Viral altruism? charitable giving and social contagion in online networks. *Sociological Science*, 3, 202–238.
- Lee, Y.-K., & Chang, C.-T. (2007). Who gives what to charity? characteristics affecting donation behavior. *Social Behavior and Personality*, 35(9), 1173–1180.
- List, J. A., & Lucking-Reiley, D. (2002). The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign. *Journal of Political Economy*, 110(1), 215–233.
- Macdonald, J. S., & Macdonald, L. D. (1964). Chain migration, ethnic neighborhood formation and social networks. *The Milbank Memorial Fund Quarterly*, 42(1), 82–97. Retrieved from <http://www.jstor.org/stable/3348581>
- Mermelstein, R., Cohen, S., Lichtenstein, E., Baer, J. S., & Kamarck, T. (1986). Social support and smoking cessation and maintenance. *Journal of consulting and clinical psychology*, 54(4), 447.
- Morris, S. (2000). Contagion. *Review of Economic Studies*, 67(1), 57–78. doi: 10.1111/1467-937X.00121
- Mossel, E., & Roch, S. (2010). Submodularity of influence in social networks: From local to global. *SIAM J. Comput.*, 39(6), 2176–2188.
- Myneni, S., Cobb, N. K., & Cohen, T. (2013). Finding meaning in social media: content-based social network analysis of quitnet to identify new opportunities for health promotion. In *Medinfo* (pp. 807–811).
- Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on world wide web* (pp. 695–704).

Ugander, J., Backstrom, L., Marlow, C., & Kleinberg, J. (2012, April). Structural diversity in social contagion. *Proc. National Academy of Sciences*, 109(16), 5962–5966.

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9), 5766–5771. Retrieved from <http://www.pnas.org/content/99/9/5766.abstract>

Table 1

Experiment A: The number of donations received for each charity.

	Charity Names	Freq	%
1	Keep 10 cents and Do Not Donate	885	61.7%
2	Project Healthy Children	171	11.9%
3	Centre for HIV and AIDS Prevention Studies	65	4.5%
4	Global Fund to Fight AIDS	53	3.7%
5	Against Malaria Foundation	41	3.9%
6	Food Fortification Initiative	37	2.6%
7	Global Alliance for Improved Nutrition	35	2.4%
8	Schistosomiasis Control Initiative	30	2.1%
9	Sightsavers	20	1.4%
10	Meningitis Vaccine Project	19	1.3%
11	GiveDirectly.org	18	1.2%
12	Living Goods	16	1.1%
13	Development Media International	13	0.9%
14	Deworm the World Initiative led by Evidence Action	13	0.9%
15	Measles and Rubella Initiative	13	0.9%
16	END Fund	6	0.4%

Table 2

Experiment A: Logistic regression for donation to any charity.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.8543	0.1698	-5.03	0.0000
DonatingNeighbors	0.1624	0.0737	2.20	0.0275
SimilarityLow	-0.0798	0.2411	-0.33	0.7406
DonatingNeighbors:SimilarityLow	0.0876	0.1044	0.84	0.4015

Table 3

Experiment A: Logistic regression for donation to SCI.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.1888	0.6487	-8.00	0.0000
SchisNeighbors	0.8135	0.2918	2.79	0.0053
SimilarityLow	-0.1483	0.9515	-0.16	0.8762
SchisNeighbors:SimilarityLow	-0.0001	0.4270	-0.00	0.9998

Table 4

Experiment B: The number of donations received for each charity.

	Charity Names	Freq	%
1	Keep 10 cents and Do Not Donate	774	53.6%
2	Project Healthy Children	185	12.8%
3	Against Malaria Foundation	70	4.9%
4	Centre for HIV and AIDS Prevention Studies	65	4.5%
5	Schistosomiasis Control Initiative	59	4.1%
6	Food Fortification Initiative	52	3.6%
7	Global Fund to Fight AIDS	52	3.6%
8	Global Alliance for Improved Nutrition	40	2.8%
9	Living Goods	37	2.6%
10	GiveDirectly.org	26	1.8%
11	Sightsavers	25	1.7%
12	Meningitis Vaccine Project	17	1.2%
13	Development Media International	17	1.2%
14	Measles and Rubella Initiative	11	0.8%
15	Deworm the World Initiative led by Evidence Action	9	0.6%
16	END Fund	4	0.3%

Table 5

Experiment B: Logistic regression for donation to any charity.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.7210	0.1642	-4.39	0.0000
DonatingNeighbors	0.2706	0.0718	3.77	0.0002
SimilarityLow	-0.0151	0.2339	-0.06	0.9484
DonatingNeighbors:SimilarityLow	0.0367	0.1020	0.36	0.7188

Table 6

Experiment B: Logistic regression for donation to SCI.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.8528	0.5059	-9.59	0.0000
SchisNeighbors	0.9832	0.2217	4.44	<0.0001
SimilarityLow	-0.1439	0.7388	-0.19	0.8456
SchisNeighbors:SimilarityLow	0.0131	0.3224	0.04	0.9675

Table 7

Experiment C: Logistic regression for donation to SCI.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.7916	0.7573	-7.65	0.0000
SchisNeighbors	1.0490	0.2908	3.61	0.0003
SimilarityLow	0.7766	0.9586	0.81	0.4179
SchisNeighbors:SimilarityLow	-0.3162	0.3769	-0.84	0.4015

Figure 1. Submodular vs. non-submodular influence curve

Groups forming...

- You are being assigned to a group of 4 users.
- You and your groupmates will be offered the chance to donate 10 cents of your earnings to charity.
- Your groupmates might or might not have answered the 10 questions similarly to you.
- **High Similarity:** More than 7 answers in common with you.
- **Medium Similarity:** Between 3 and 6 answers in common with you.
- **Low Similarity:** Less than 3 answers in common with you.

User	Pic	Similarity to You	Can See Charity Choice of
You		--
A		Low	Nobody
B		Low	Nobody

Figure 2. Phase 2: Group Formation. The subject viewed a dynamic display that added neighbors (groupmates) one at a time.

Where would you like to donate?

We have been asking other members of your group to donate 10 cents of their earnings to the charity of their choice.

A

Participant A has decided:
Sightsavers

B

Participant B has decided:
Schistosomiasis Control Initiative

C

Participant C has decided:
Schistosomiasis Control Initiative

Would you also like to donate 10 cents of your earnings to the charity of your choice?

☐ Against Malaria Foundation	<div><div><div></div><div></div></div>☐ Schistosomiasis Control Initiative</div>	☐ Deworm the World Initiative led by Evidence Action	☐ Measles and Rubella Initiative
☐ Living Goods	☐ Development Media International	☐ Food Fortification Initiative	☐ Meningitis Vaccine Project
☐ Global Alliance for Improved Nutrition	☐ Project Healthy Children	☐ Global Fund to Fight AIDS	☐ Centre for HIV and AIDS Prevention Studies
☐ END Fund	<div><div><div></div><div></div></div>☐ Sightsavers</div>	☐ GiveDirectly.org	☐ Keep 10 cents and Do Not Donate

Figure 3. Phase 3: Donation Choice. The subject chooses where to allocate their earnings. Groupmates’ choices populate as the screen is viewed.

Survey and Demographic Questions

Please provide the following demographic information.

Gender ☐ Female ☐ Male

Please select the annual income range you belong to: ☐ Under \$25000 ☐ between \$25000 and \$60000 ☐ More than \$60000

Year of Birth

Country of residence

Is English your first language? ☐ Yes ☐ No

Have you ever tried smoking? ☐ Yes ☐ No

Figure 4. Phase 4: Survey and Demographic Questions. Subjects were prompted for demographic information after completing the experiment.

Figure 5. Top: Proportion of participants donating to any charity versus # neighbors donating to any charity. Bottom: Proportion participants donating to SCI versus # neighbors donating to SCI.

Figure 6. Proportion of participants matching neighbor behavior as a function of neighbor count and similarity. There is no evidence similarity affects influence by interacting with neighbor count.

Figure 7. Experiment B Influence Curves. Top: Proportion of participants donating to any charity versus # neighbors donating to any charity. Bottom: % participants donating to SCI versus # neighbors donating to SCI.

Figure 8. Experiment C: Proportion participants donating to SCI versus # neighbors donating to SCI.

Appendix

Supplemental Information

Distribution of Subjects over Conditions for Experiment A

Table A1

The number of participants per condition in Experiment A. The first segment describes the Similarity condition - groupmates were either High Similarity (AllHi_) or Low Similarity (AllLo_). The next segment (Self_) precedes the number of neighbors (bots) choosing to Keep the 10 Cents. The next segment (Schis_) precedes the number of neighbors (bots) choosing to donate to the Schistosomiasis Control Initiative. The last segment (Rand_) precedes the number of neighbors (bots) choosing to donate to a charity other than Schistosomiasis Control Initiative (randomly chosen without replacement). All neighbor counts sum to 3.

	Condition	Freq
1	AllHi__Self_0__Schis_1__Rand_2	108
2	AllHi__Self_0__Schis_2__Rand_1	100
3	AllHi__Self_0__Schis_3__Rand_0	100
4	AllHi__Self_1__Schis_1__Rand_1	102
5	AllHi__Self_1__Schis_2__Rand_0	101
6	AllHi__Self_2__Schis_1__Rand_0	100
7	AllHi__Self_3__Schis_0__Rand_0	105
8	AllLo__Self_0__Schis_1__Rand_2	106
9	AllLo__Self_0__Schis_2__Rand_1	101
10	AllLo__Self_0__Schis_3__Rand_0	102
11	AllLo__Self_1__Schis_1__Rand_1	101
12	AllLo__Self_1__Schis_2__Rand_0	101
13	AllLo__Self_2__Schis_1__Rand_0	105
14	AllLo__Self_3__Schis_0__Rand_0	103

Questions from Experiment A

The screenshot shows a web interface titled "Question Phase" in a large, bold, black font. Below the title, there are two bullet points: "Please answer the following 10 questions." and "After you are done with your responses, click the Continue button." Below these instructions, there are two separate question boxes. The first box contains the question "Are you a cat person or a dog person?" with two radio button options: "Cat" and "Dog". The second box contains the question "Do you prefer outdoor activities or indoor activities?" with two radio button options: "Outdoor" and "Indoor".

Question Phase

- Please answer the following 10 questions.
- After you are done with your responses, click the Continue button.

Are you a cat person or a dog person?

☐ Cat

☐ Dog

Do you prefer outdoor activities or indoor activities?

☐ Outdoor

☐ Indoor

Figure A1. Phase 1: Personal Preference Questions. This screenshot displays two questions as the subjects would have seen them in their web browser.

The subjects were asked to answer the following ten questions.

1. Are you a cat person or a dog person?
 - Cat
 - Dog
2. Do you prefer outdoor activities or indoor activities?
 - Outdoor
 - Indoor
3. Do you prefer living in a big city or in the countryside?

- City
- Countryside

4. Are you an introvert or an extrovert?

- Introvert
- Extrovert

5. Do you prefer playing team sports or individual sports?

- Team Sports
- Individual Sports

6. Do you prefer books or movies?

- Books
- Movies

7. Do you like scary movies or not?

- Yes
- No

8. Do you prefer cold weather or hot weather?

- Cold
- Hot

9. Do you live near the ocean (West Coast or East Coast) or more toward the middle of the country?

- Coast
- Middle

10. What types of books do you like more - fantasy/scifi books or mysteries?

- Fantasy/Science-Fiction
- Mysteries

Politically Polarizing Questions from Experiments B and C

1. Do you believe the government should do more to restrict gun ownership?
 - Yes
 - No
2. Do you believe that climate change is more likely to be a real threat or more likely to be not a big deal?
 - Real threat
 - Not a big deal
3. Which of the following statements comes closest to your view?
 - Racial discrimination is the main reason why many black people can't get ahead these days
 - Blacks who can't get ahead in this country are mostly responsible for their own condition
4. Which of the following statements comes closest to your view?
 - Homosexuality should be accepted by society
 - Homosexuality should be discouraged by society
5. Which of the following statements comes closest to your view?
 - The obstacles that once made it harder for women than men to get ahead are now largely gone
 - There are still significant obstacles that make it harder for women to get ahead than men

6. Do you believe that abortion is murder?

- Yes
- No

7. Should Muslim immigrants be allowed in the US?

- Yes
- No

8. Do you believe the United States would be better if it were a more Christian nation?

- Yes
- No

9. Should public schools be allowed to start the day with a student-led prayer?

- Yes
- No

10. Did Russians influence the outcome of the 2016 US Presidential Election?

- Yes
- No

The Chen-Szroeter Test

Chen and Szroeter Chen and Szroeter (2009) develop a natural, computationally efficient procedure that takes in multiple tests and outputs the p -value of the event that at least one of the tests is violated. To be precise, the null hypothesis is that none the tests are violated. The p -value then represents a bound on the probability of the data, given that none the tests are violated. There procedure explicitly takes into account the covariance of the tests. They prove that their test is asymptotically exact. That is, as the number of data become large, the p -value approaches the correct value. The test is

based on a chaining argument, which we now briefly describe. Let $\mu_1, \dots, \mu_\ell$ be a set of parameters in a model. The null hypothesis H_0 states that $\mu_i \geq 0$ for all $1 \leq i \leq \ell$. The complement hypothesis H_1 states that there exists an $1 \leq i \leq \ell$ such that $\mu_i < 0$. Let $D(\cdot) : \mathbf{R} \rightarrow \mathbf{R}$ be the indicator function such that $D(x) = 1$ if $x < 0$ and $D(x) = 0$ otherwise. We can then test $H_0 : \sum_{j=1}^{\ell} D(\mu_j)\mu_j = 0$ versus $H_1 : \sum_{j=1}^{\ell} D(\mu_j)\mu_j < 0$. An issue is that D has a large discontinuity at 0, but by replacing D with a series of smoothed functions that point-wise converge to D the result is obtained.

Before we talk further about combining the tests, let's recall the individual tests that should be satisfied.

Let $P(i)$ be the probability of donating conditioned on being in a treatment were i of your neighbors donate. The if the adoption curve is submodular, 5 different constraints should be satisfied:

$$P(2) - P(1) \leq P(1) - P(0)$$

$$P(3) - P(2) \leq P(2) - P(1)$$

$$P(3) - P(2) \leq P(1) - P(0)$$

$$P(3) - P(1) \leq 2(P(1) - P(0))$$

$$2(P(3) - P(2)) \leq P(2) - P(0)$$

If even one of these fails, then the function is non-submodular. We can easily rewrite these in the form required by Chen and Szroeter's chaining method. For example, let $\mu_1 = P(1) - P(0) - P(2) + P(1)$. Then $\mu_1 \geq 0$ exactly when $P(2) - P(1) \leq P(1) - P(0)$. This μ_i expression is often called a contrast. Here, the contrasts are always in a special form called the 'difference of differences'.

Given that the subjects are assumed to be independent, we have binomial estimates $\hat{P}(i)$ for the $P(i)$. Thus, in the limit, the standard deviation of each contrast μ_i is easy to estimate. For example, if we would like to test the contrast $C = P(i_1) - P(i_2) - P(i_3) + P(i_4) > 0$. The variance of the contrast expression, when i_1, i_2, i_3, i_4 are all distinct, is

$$V = \sum_{k=1}^4 \frac{P(i_k)(1 - P(i_k))}{n_{i_k}}$$

where n_{i_k} is the sample size of each treatment.

If we want to test whether $\mu_i > 0$ individually, we form a test statistic by taking the contrast divided by the standard deviation, i.e. $C/\sqrt{V}$. If the population contrast is zero then this is approximately standard normal. We use a normal probability table to construct a p -value or confidence interval (CI). Note that if any of the variables i_1, i_2, i_3, i_4 are the same, then we need to recalculate the variance. For example, the variance of $C = P(i_1) - 2P(i_2) + P(i_4) > 0$ would be

$$\frac{P(i_1)(1 - P(i_2))}{n_{i_1}} + \frac{P(i_2)(1 - P(i_2))}{4n_{i_2}} + \frac{P(i_3)(1 - P(i_3))}{n_{i_3}}.$$

Using the Chen and Szroeter test, we can test if any of these constraints is violated. We only need to compute the covariance between the test statistics, which is just the sum of the terms $\frac{P(i_k)(1-P(i_k))}{n_{i_k} c_{i_1} c_{i_2}}$ that the tests have in common where c_{i_1} and c_{i_2} are just the coefficients of the terms in each contrast.

The full details of how we implement the test are in section 3 of Chen and Szroeter (2009). We note that there are four different flavors of the Chen test depending on what type of correction is used for multiple tests (either linear or quadratic), and what type of error function is used around 0 (either a threshold or soft-max). We find that these parameters did not make a noticeable difference, and simply report the first ones mentioned in the paper—linear and threshold. There is also a scaling parameter, which we conservatively set to the size of the smallest sample.

Calibrating the Chen and Szroeter Test. Formally, the paper shows that this test is asymptotically accurate. That is, the test correctly outputs the probability of the results conditioned on the null hypothesis being true (all tests are satisfied by the model).

However, our data is not asymptotically large, so we ran a battery of tests on finite data size, and found that the test performed surprisingly well for p -values in the range of 0.05 and below. However, for larger p -values (e.g. 0.2), while the test performed accurately for a large number of samples, the test was occasionally optimistic on data from fewer samples. To test the method, we simulated our experiment being run when for different values of the $P(i)$'s. First, we found that, if the $P(i)$'s were

strictly non-submodular, the p -value would approach 1 as the number of simulated samples in each treatment grew. If the $P(i)$'s were strictly non-submodular, the p -value would approach 0.

One interesting case is where the $P(i)$ s are linear. For example $P(i) = .1 + i * .1$. This is because, here all the test are just barely satisfied. In this case, we found that as the number of samples increased, the probability of getting a p -values less than .05 was strictly less than .05. For example, with the above $P(i)$ s using 10,000 samples in each treatment, out of 10,000 runs⁴, only a 0.0201 fraction had p -values less than 0.05. Moreover, for smaller samples, the probability of getting a p -value of less than .05 was consistently around the same size or smaller than for larger samples. For example, using the size of the first experiment with the above $P(i)$ s the probability of getting a p -value of less than .05 was 0.0204.

We tried one more type of test, where exactly one constraint was tight. For example $P(0) = .2$, $P(1) = .4$, $P(2) = .6$ and $P(3) = .7$. Note here that $P(2) - P(1) = P(1) - P(0)$. In these cases, we found that the the fraction of simulations with a p -value of less than 0.5 was approximately .0481 when simulating 10,000 samples in each treatment. When using the smaller sample that was actually run in the test, this decreases to a 0.0062 fraction of runs having p -value at most 0.05.

⁴ we always use 10,000 samples and 10,000 runs unless we explicitly report otherwise