Polit Behav
DOI 10.1007/s11109-015-9308-6

ORIGINAL PAPER

When Treatments are Tweets: A Network Mobilization Experiment over Twitter

Alexander Coppock[1] · Andrew Guess[1] · John Ternovski[2]

Abstract This study rigorously compares the effectiveness of online mobilization appeals via two randomized field experiments conducted over the social microblogging service Twitter. In the process, we demonstrate a methodological innovation designed to capture social effects by exogenously inducing network behavior. In both experiments, we find that direct, private messages to followers of a nonprofit advocacy organization's Twitter account are highly effective at increasing support for an online petition. Surprisingly, public tweets have no effect at all. We additionally randomize the private messages to prime subjects with either a "follower" or an "organizer" identity but find no evidence that this affects the likelihood of signing the petition. Finally, in the second experiment, followers of subjects induced to tweet a link to the petition are more likely to sign it—evidence of a campaign gone "viral." In presenting these results, we contribute to a nascent body of experimental literature exploring political behavior in online social media.

Keywords Field experiments · Networks · Political participation

Electronic supplementary material The online version of this article (doi:10.1007/s11109-015-9308-6) contains supplementary material, which is available to authorized users.

✉ Andrew Guess
a.guess@columbia.edu

1 Department of Political Science, Columbia University, 420 W. 118th St., Suite 710, New York, NY 10027, USA

2 Oxford Internet Institute, University of Oxford, Oxford, UK

Published online: 08 April 2015

Introduction

While much enthusiasm about the Internet focuses on its ability to foster informal and decentralized forms of organization (Shirky 2008; Benkler 2006; Bennett and Segerberg 2012), traditional groups have long recognized its potential for recruitment and mobilization (Obar et al. 2012). This is especially true in the realm of politics: With meaningful political behavior now commonplace online, campaigns have added email and Facebook appeals to their arsenal of tactics (Krueger 2006; Gaby and Caren 2012). Nonpartisan and advocacy organizations have similarly turned to social media to engage their supporters.

This study examines the effects of an online mobilization campaign via what we believe to be the first randomized field experiments conducted on the social network Twitter. Our design allows us to identify the effects of both private messaging and "natural" network behavior directed toward supporters of a nonprofit advocacy organization, the League of Conservation Voters (LCV). The two field experiments we present here follow nearly identical designs and lead us to draw very similar conclusions about political mobilization over Twitter.

Our primary manipulation in both experiments exposes some subjects to a public tweet only and others to one of two private direct messages. The wording of the direct messages primes subjects with one of two identities based on previous research demonstrating the behavioral consequences of associating political actions with a particular self-concept (Bryan et al. 2011). With this manipulation, we investigate whether the passivity associated with a "followers" label or the higher level of commitment associated with an "organizers" label has an impact on our two outcomes: signing an online petition and tweeting (or retweeting) the petition link.

Our secondary manipulation encourages a random subset of petition signers to tweet the petition to their own followers. This has the consequence of randomly assigning the followers of petition signers to be exposed to tweets directing them to the petition. This design allows us to explore network effects while avoiding homophily concerns (McPherson et al. 2001).

We find that private direct messages on Twitter are highly effective tools for generating online petition signatures, a common advocacy goal. In addition, we find no evidence that the type of identity primed affects the likelihood of signing an online petition. However, those assigned to the "follower" condition are more likely in both experiments to tweet a link to the petition to their own followers. Results from our secondary manipulation are mixed: In our second experiment but not the first, we find evidence of network effects among followers of the organization's own followers. Finally and most surprisingly, no one who was exposed to only the public tweet either signed the petition or tweeted the link to their own followers.

In addition to providing practical guidance for organizations with dedicated follower networks, these results suggest that the advantages of personal appeals identified in face-to-face campaigns can carry over into the virtual world (Rosenstone and Hansen 1993; Gerber and Green 2000). They also show that invoking identities associated with different levels of commitment to a cause can affect one's propensity to comply with a simple request. This evidence has potential implications for the continuing debate on whether online campaigns of the type studied here merely promote "slacktivism"—token, low-cost emblems of support incapable of sustaining meaningful collective action (Morozov 2009; Gladwell 2010).

This paper proceeds as follows. In the next section, we review recent theoretical arguments on whether social media can facilitate or hinder collective action. In Sect. 3, we outline the potential pitfalls of analyzing experimental manipulations over social networks like Twitter. In Sect. 4, we provide an overview of the universe of subjects—the follower network of a large nonprofit advocacy organization—and place it in context. Section 5 describes the research design and analytic strategy of the two experiments. Sections 6 and 7 report the results of both experiments, and Sect. 8 describes the heterogeneous effects of treatment by account type. Section 9 concludes with a discussion.

Participation, Collective Action, and Social Media

Scholars have sought to understand how lowered communication costs and the proliferation of online social networks have altered the traditional logic of collective action (Olson 1965). One set of responses focuses on the scope of action: The relatively low-effort individual contributions involved in online campaigns necessarily limit the value of resulting public goods (Gladwell 2010; Shulman 2009). In contrast to this pessimistic assessment, some communication scholars—basing their insights on grassroots anti-globalization campaigns or, more recently, events leading up to the Arab Spring—delineate how new communication technologies can foster real-world spontaneous collective action in the absence of formal organization or leadership (Bennett and Segerberg 2012). These theories extend Olson's classic work to networked, often digitally mediated contexts.

Little theoretical work thus far explicitly invokes Twitter (see Marwick and boyd 2011 for an exception), but many of the ideas carry over from discussions of viral email campaigns. For example, Bimber et al. (2005) associate the act of contributing to a collective good with a transition across the boundary from private to public. However, because the new technologies often blur the line between public and private, they find that "boundary crossing in connection with public goods takes on forms not so readily recognizable in the theoretical terms of free riding, selective incentives, and organization" (p. 378). In this alternative framework, forwarding a petition (or, perhaps, retweeting a message) is a "nearly costless request" to make private or semiprivate information—that is, the fact that someone supports a campaign—public.[1]

Another perspective, championed by Benkler (2006) and Bennett and Segerberg (2012), emphasizes how networks enable the co-production of collective goods, fostered via individual self-expression. Prototypical examples in this tradition

[1] Sharing private information with third parties is now an ingrained part of online behavior, with potential benefits for both traditional advocacy groups and networked movements.

include the open-source software movement and Wikipedia, relatively decentralized networks of contributors who help to create and maintain free public resources. In the political realm, the movement against the Stop Online Piracy Act (SOPA) or for net neutrality might fall in this category of "connective action," although the type of mobilization studied here falls more clearly under the traditional umbrella of "organizationally brokered networks" (Bennett and Segerberg 2012, p. 756). As we demonstrate in these experiments, networked participation can arise from centrally organized campaigns as well.

Classic treatments of participation in politics understandably focus on factors affecting the likelihood that different types of citizens will vote, volunteer for a candidate, or contribute in some other way on behalf of a political cause. This study necessarily approaches the topic differently, focusing on people who have already self-selected as followers of a prominent environmental advocacy organization. In effect, we condition on the usual determinants of participation as presented in the Civic Voluntarism Model (Verba et al. 1995): resources, motivation, and intensive engagement with a political issue. Accordingly, our experiments focus on the model's remaining ingredient, "networks of recruitment" (p. 3).[2] Although Verba et al. assumed that such networks would consist of commonly studied social ties such as friends, family, and co-workers, we extend this notion to include connections based on affinity.

Recruitment networks are a useful concept for studying mobilization in an online context. Aside from capturing the possibility of "viral" or network patterns in campaign activity, they allow us to compare the effectiveness of appeals that originate from the organization itself with those from non-affiliated but like-minded supporters of its environmental mission. This distinction between direct appeals and peer effects is important for at least two reasons. The first relates to Olson's original observation about the "noticeability" of individual contributions, which implies that as organizations become large, shirking becomes unobservable to other members and free-riding inevitable (barring selective incentives, coercion, or other inducements to cooperate). Peer effects facilitated via transparent social networks, by contrast, are one possible way in which online organizing could overcome the problem of "noticeability" and mitigate the incentives to free-ride in large groups (Lupia and Sin 2003).

Second, the distinction matters because social ties may reflect group membership. This insight arises from the Elaborated Social Identity Model, which was constructed to explain how group membership can induce collective action (Drury et al. 2005). The theory presupposes the existence of a grassroots in-group and a powerful out-group; any action taken by the in-group against the out-group that appears to succeed is "experienced as joyful and exhilarating" (Barr and Drury 2009, p. 245). This model has been applied to induce voter mobilization by labeling targets as "voters" rather than simply people who vote (Bryan et al. 2011). "Donor" and "activist" identities have also been associated with increased

[2] The organization-centered design also addresses any concern about our lack of covariates for these traditional predictors of participation, although Sect. 8 analyzes differential effects by factors that we were able to capture, gender and organizational status.

charitable donations and activism, respectively (Aaker and Akutsu 2009), though no published large-scale field experiments have evaluated the impact of the "donor," "organizer", or "activist" identity labels on those outcomes.

With our design, elaborated in Sect. 5, we simultaneously address several of the theoretical debates raised in the literature. First, we measure the effects of our manipulations on two primary outcomes: filling out (or "signing") an online petition, and tweeting (or retweeting) a link to the petition to one's own followers. The former is recognizable as a contribution to a public good, one traditionally valued as part of the political organizer's toolbox (Karpf 2010). The latter is a somewhat more ambiguous—but arguably less costly—action that, if repeated by many other members, could lead to increased public awareness of the campaign (and its magnitude of support). An important difference between the two outcomes is that tweeting more directly captures whether the "noticeability" of the behavior leads to enhanced effectiveness, allowing for a test of how Olson's logic may operate differently when mediated via online communication networks. Second, we examine the effects of three types of appeals: Generic appeals via the organization's public Twitter account, specific appeals via private direct messages, and tweets from followers of the organization to their own followers. These differ in the extent to which they depend on social ties, direct contact, and the authority of a trusted organization. Finally, we vary the identity labels used to address our subjects, enabling us to examine whether the salience of specific social identities is associated with the likelihood of contributing to the organization's goals.

To briefly summarize the expectations of the participation and online collective action literature to date, we believe that the work of Bimber et al. (2005) would lead to a prediction of larger treatment effects for the (re)tweeting outcome, which involves the relatively low-cost act of making one's support for a cause (more) public. The "slacktivism" hypothesis, by contrast, straightforwardly predicts strong treatment effects for the lowest-cost actions regardless of whether they make information public. This leads to the expectation that subjects across all conditions will retweet LCV's public tweet but that fewer will take the time to sign the petition (and subsequently tweet out the link to it). Finally, the Civic Voluntarism Model (Verba et al. 1995) predicts that "networks of recruitment" will be most effective at mobilizing followers: The largest treatment effects will be observed when subjects are exposed to tweets from peers.

The Challenges of Experiments on Twitter

In the terminology of network analysis, the social microblogging service Twitter is a directed graph. Users post short, public updates and curate their own networks by "following" others—friends and strangers alike—who may or may not reciprocate. A particular user's Twitter messages, or "tweets", can be read by anyone who visits his or her public feed (also known as a timeline). Since manually reading individual feeds can be cumbersome, users typically take advantage of the Twitter "stream," a real-time aggregation of tweets from users they follow. The result is a never-ending rush of text and photo updates from sources of a user's choosing.

There are several specific ways of communicating on Twitter. Most fundamental is the tweet, usually restricted to 140 characters (with exceptions for web addresses of reasonable length). "Retweets" (RTs) allow users to quickly resend a tweet from their stream to those of their own followers (in other words, a retweet copies a tweet from a user's incoming stream to his or her own feed, with attribution). In extreme cases, this capability can lead to cascades of retweets of particularly compelling content. Other tweets, while not necessarily retweets, can "mention" another user. This option can lead to extended public conversations, all potentially referring back to an initial Twitter posting. Finally, while Twitter is best known for its public functionality, it also allows users to send private "direct messages" (DMs) to any of their followers. By default, Twitter sends users an email notification when they receive a DM.

Public tweets comprise the bulk of a typical Twitter account's activity, yet they present challenges for the design and analysis of experiments. Since public tweets and retweets are visible to followers of the sender, a simple experimental design in which some followers are randomized to be shown a public tweet and others in a control group are not is effectively impossible. An alternative design would randomly time a series of public tweets with different messages, but any causal inferences would require strong modeling assumptions concerning the over-time persistence of treatment effects (for example, if the same message is tweeted every other day, followers may become irritated and respond differently from how they otherwise would).

Instead, we take advantage of Twitter's direct message capability, which allows us to present different messages to different users. Estimation of the relative effectiveness of the messages can proceed in the normal fashion, under an assumption of non-interference between units. The non-interference assumption (sometimes referred to as the stable unit treatment value assumption, or SUTVA) requires that subjects' outcomes not be influenced by the treatment assignments of other subjects. The interference concern is not trivial: unmodeled spillovers can lead to biased estimates of treatment effects (Gerber and Green 2012, Chapt. 8). For example, if direct messages were highly effective at motivating petition signatures and subsequent tweets, but those tweets exposed subjects in the control group to the same message, a naive difference-in-means estimate would be biased.

A schematic version of an analytic approach to dealing with spillovers of this kind is as follows: First, redefine treatment categories to include "spillover conditions" such as being in the condition of following one subject who received a direct message. Second, calculate the probability that each unit is in each redefined treatment condition. Because Twitter users follow vastly different numbers of other users, these probabilities will vary quite a bit from unit to unit. Third, weight each unit's outcome by the inverse of the probability of being in its observed condition. Average differences across these redefined treatment categories will reflect unbiased treatment effect estimates.

The trouble with this approach is the prohibitively large number of potential treatment categories: anywhere from following zero treated units to following 601 (the largest out-degree observed in our network). One could instead parameterize the "dosage" of spillovers and estimate a response curve for each extra treated unit (see Bowers et al. 2013 for the method, Coppock 2014 for an application). Such a method, however, requires the researcher to make strong functional form assumptions concerning exposure.

Our solution is to vastly reduce the number of potential spillover conditions by design. As discussed in Sect. 5 below, we randomly induce a relatively small subset of the network to retweet to their followers, which in turn randomly exposes some users but not others to the petition link.[3] Most users only follow one user in this subset, so the number of spillover conditions is quite manageable. Additionally, the variability of exposure probabilities is kept in check. In effect, this approach allows us to randomize "natural" Twitter behavior and directly estimate its consequences.

Because of the difficulties outlined here, it is not surprising that randomized experiments on Twitter have been rare. One study randomly encouraged subjects to follow a Japanese politician on Twitter to test effects on trait evaluations, knowledge, and other post-treatment outcomes (Kobayashi and Ichifuji 2014). A marketing study tested the effectiveness of tweets and retweets on television ratings by collaborating with a media company, as well as "influential" tweeters, to promote a random subset of its TV shows on the Chinese social network Sina Weibo (Gong et al. 2014). The unit of analysis in this design was the shows themselves. Finally, there is ongoing research on the relationship between identity and behavior on social media. A randomized experiment on a web-based social sharing site (but not Twitter) found that cues indicating an account's identity matter in terms of how users share content associated with that account (Taylor et al. 2014).

Overview of Network

The League of Conservation Voters is an environmental advocacy organization that "works to turn environmental values into national, state and local priorities," according to its official website. Its activities include public awareness campaigns, lobbying efforts, and independent expenditures (via political action committees) geared toward electing candidates who support its agenda.

The two field experiments analyzed here were deployed over the network of followers of LCV's official Twitter account. The members of this network comprise an "issue public" in the literal sense (Converse 1964; Verba et al. 1995): highly dedicated to promoting the environment and publicly visible in their activism. Studying the impact of mobilization tactics on such a network speaks to the effectiveness of promoting activism among politically engaged individuals, although any generalizations to other populations would have to be qualified.

While the network is highly engaged, it is not a close-knit community. Rather, it is mainly a network of strangers. Members all share a common interest in the environment, but there are relatively few interconnections between them: out of a possible 44,709,282 connections between nodes, there were 131,474 such edges

[3] Contrast this with the case in which the organization is sending the public tweets—all of their followers are potentially exposed. When a random subset of users tweet the link, however, only a portion of the organization's followers are exposed, allowing for experimental differences to be observed.

when the network was scraped before Study 1, yielding a graph density of 0.0029. Put more simply, followers of the organization follow a median of only six users in the organization's network. This structure corresponds most closely to a "Broadcast Network," as described by the Pew Research Center's typology of conversational archetypes on Twitter.[4] Such "hub and spoke" networks consist of an audience of followers who typically rebroadcast (i.e., retweet) the output of a single source, in this case LCV.

An additional feature of the network, as Fig. 1 illustrates, is its relatively diffuse nature. Unlike a highly modular social network with various distinct groupings, this one has numerous and overlapping communities that are difficult to distinguish from each other. Network statistics confirm this impression: The Walktrap algorithm (with standard defaults) finds 22 communities in the network, for a modularity of 0.3.

Describing a network with statistics such as the graph density and the modularity is usually insufficient for conveying its structure. Modularity, for example, depends on the number of communities detected; different algorithms come to different conclusions about the number of communities and their membership. In our view, the principal utility of these statistics is to provide partial justification for our description of the subject pool as a being a social community only in the loosest sense.

For Study 1, we constructed our universe of subjects by scraping the Twitter ID numbers of LCV's followers, excluding those who had more than 5000 followers of their own. The reasoning behind this decision was that users with especially large numbers of followers were more likely to be prominent individuals or organizations whose online behavior would not be comparable with the rest of the subject pool. The resulting network contained 6687 members. For Study 2, conducted five months later, we repeated this procedure and obtained a network with 8507 members.

LCV's Twitter account is fairly active, sending out an average of 6.07 tweets and retweets per weekday from February 2013 to February 2015. However, as Fig. 2 shows, day-to-day variation in the number of tweets posted is high (s.d. = 10.05). This activity has not stopped the network from continuing to grow, suggesting that its followers, in addition to being dedicated to the cause, are accustomed to frequent Twitter updates from the organization.

This is likely also the case for comparable organizations. Table 1 lists the number of followers and average number of tweets per weekday for the top ten most influential environmental organizations (by 2014 lobbying expenditures as collected by OpenSecrets.org). The mean number of followers among this group, 118,988, is an order of magnitude higher than LCV's number of followers, and average tweet frequency—about 13 per weekday—is slightly more than twice as high. Finally, the median creation date for these organizations' Twitter accounts was mid-2008, fairly early in Twitter's history.

[4] See http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/.

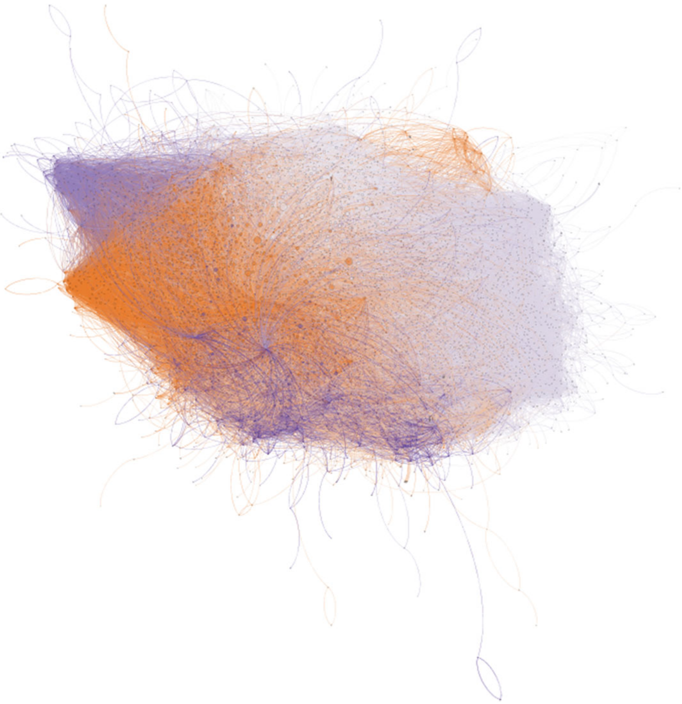

Fig. 1 An illustration of the network of 6687 followers of the advocacy organization's Twitter account scraped before Study 1. Lines illustrate connections between users and are shaded by membership in one of 22 communities as determined by the Walktrap community detection algorithm

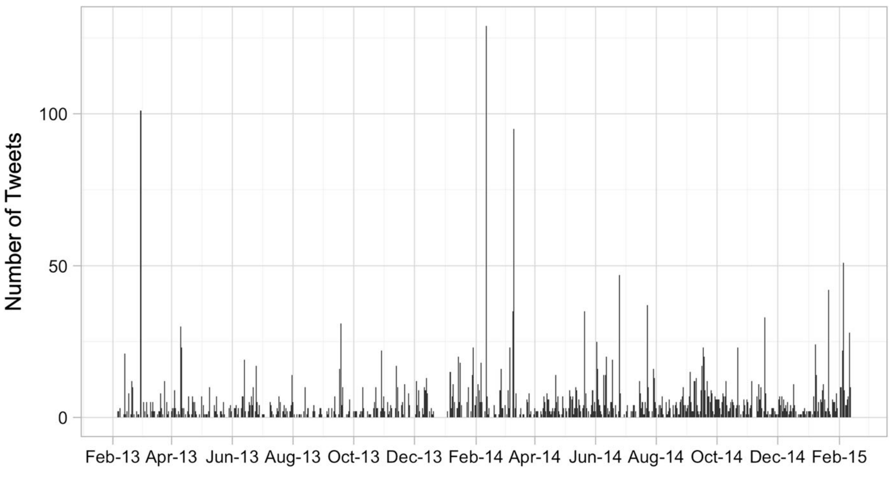

Fig. 2 The number of tweets and retweets sent by LCV per day, from February 2013 to February 2015

Table 1 Descriptive statistics (circa February 2015) of Twitter accounts of the top ten environmental organizations by 2014 lobbying expenditures, plus LCV in bold

Organization	Account	# Followers	# Tweets/ weekday	Created
Earthjustice Legal Defense Fund	@Earthjustice	67,004	8.60	2008
Environmental Defense Fund	@EnvDefenseFund	93,619	6.87	2009
Environmental Working Group	@ewg	35,279	7.65	2008
Intl Assn of Fish & Wildlife Agencies	@fishwildlife	1229	1.30	2010
League of Conservation Voters	**@LCVoters**	**10,880**	**6.07**	**2009**
Nature Conservancy	@nature_org	387,973	9.60	2008
National Parks Conservation Assn	@NPCA	108,164	11.47	2009
Natural Resources Defense Council	@NRDCFedGov	681	2.10	2012
National Wildlife Federation	@NWF	296,513	14.60	2007
Sierra Club	@sierraclub	143,448	42.67	2009
Wilderness Society	@Wilderness	55,971	28.07	2007

The number of tweets per weekday counts retweets

Research Design

We conducted two field experiments with nearly identical designs; lessons learned from the first experiment improved the design of the second. In both studies, LCV first posted a public tweet urging supporters to sign an online petition and retweet the link to their own followers. In a first-stage experiment, subjects were randomly assigned to one of three groups: (1) the baseline or control group, which was exposed to the public tweet only; (2) a condition in which subjects also received a DM with a similar request, referring to them as “followers”; (3) a condition in which subjects also received a DM referring to them as “organizers.” In a second-stage experiment, those who completed the petition were randomly shown a link with an encouragement to tweet the petition to their own followers. We will refer to this treatment as the “tweet encouragement” or the “tweet link.”

Study 1

In Study 1, LCV’s tweet and petition were related to an ongoing campaign to end tax breaks to “Big Oil”. The public tweet, posted on February 5, 2014, was followed by the DMs,[5] which had to be sent in 12 daily batches.[6] (See Online Appendix 4 for the full text of all messages.) In this version of the design, subjects

[5] “Follower” condition: “You’re one of our most valuable followers! Please RT this petition to your friends to stop tax breaks to Big Oil (URL to petition)”; “organizer” petition: “You’re one of our most valuable organizers! Please RT this petition to your friends to stop tax breaks to Big Oil (URL to petition)”

[6] Twitter’s API limits the number of DMs that an application can send to 250 per day. Subjects were randomly selected into batches.

who completed the petition were required to enter their Twitter usernames in order to connect responses back to treatment assignment.

The initial randomization procedure assigned one-third of the subjects in each of the DM conditions to be shown a tweet encouragement after submitting the petition signature, using complete random assignment.[7] In total, subjects could be assigned to one of five treatment conditions. We collected outcome data in two concurrent ways: online petition signatures were collected by survey software, while tweet and retweet behavior was captured by scraping programs that we ran continually for the duration of the experiment. We ran multiple scrapers in parallel to guard against accidental data loss.[8]

Table 2 shows the number of subjects within each condition and the proportions signing the petition and tweeting the petition link to their followers. Table 2 also reveals an anomalous finding that suggests a potential issue with the randomization: assignment to be shown the tweet encouragement predicted petition signatures ($p < .01$). Since the encouragement was only displayed to subjects *after* signing the petition, it is possible that randomization failed to eliminate unobserved differences between subjects assigned and not assigned to the tweet encouragement condition. We exhaustively investigated the possible sources of this imbalance, such as day-of-week effects, faulty randomization procedure, and data problems, but we were unable to conclusively pinpoint the cause. The most plausible explanation is that an imbalance occurred simply by chance. In Study 2, we addressed this problem by waiting until subjects clicked through to the online petition to conduct the second-stage random assignment.

In our analysis of the first-stage experiment (the direct message treatments), we will first examine average differences across treatment assignments, with and without covariate adjustment. Using information from users' public Twitter profiles, we were able to gather the following covariates: account type (male, female, organization, or unknown), number of followers, and the number of days the account was open.[9] We also calculated each subject's eigenvector centrality, a measure of how well-connected the user is within the LCV network. For our analysis of the first-stage experiment, we will rely on a strict assumption of non-interference among units. Table 2 contains some indication that this assumption is not wholly unwarranted: Despite 65 tweets of the link throughout the network, a grand total of zero subjects in the public tweet condition signed the petition. At least among this subset, we can be sure of the non-interference assumption (see also Sinclair et al. 2012 for evidence that the non-interference assumption is well-justified in get-out-the-vote mail experiments).

The subjects in the second-stage experiment (the encouragement to retweet) were those who met the following criteria: they followed both LCV and users who signed

[7] This was done by using two different versions of the petition. See Fig. A1 in Online Appendix 4 for a screen shot of the encouragement, and Fig. A2 for the tweet window that popped up if a user clicked on the tweet encouragement.

[8] By "scraping" we mean continually querying the Twitter API in order to capture tweets containing the URLs to either version of the petition. We used this approach because Twitter data is much easier to collect in real time than after the fact.

[9] See Sect. 8 for the details of this procedure.

Table 2 Study 1: design and outcomes

Treatment group	N	Signed (%)	Tweeted (%)
Public tweet	3687	0 (0.0)	0 (0.0)
Organizer DM			
Tweet encouragement	500	22 (4.4)	12 (2.4)
No encouragement	1000	28 (2.8)	12 (1.2)
Follower DM			
Tweet encouragement	500	28 (5.6)	25 (5.0)
No encouragement	1000	31 (3.1)	16 (1.6)
Total	6687	109 (1.6)	65 (1.0)

the petition. The petition received 109 signatures; these 109 users were followed by a total of 1176 other LCV followers. The 1176 were the pool of subjects randomly assigned to the condition of following someone exposed to the tweet encouragement. Similar to the procedure described in Sect. 3, we will weight each observation by the probability of exposure, as those who follow more of the 109 are more likely to be exposed. We will estimate both the intent-to-treat (ITT) effect using ordinary least squares (OLS) and the complier average causal effect (CACE) using instrumental variables (IV). The definition of a complier in the second-stage experiment is a mouthful: a complier is a user who follows one or more petition signers who tweeted the link if and only if shown the tweet encouragement. Though it may seem counterintuitive, the analysis of the second-stage experiment also assumes non-interference—we assume that a unit's outcomes do not depend on whether or not some other unit follows a petition signer assigned to the tweet encouragement.

One may wonder about users who do *not* follow LCV but may still have been exposed to tweets by virtue of following a petition signer. As it happens, exactly zero petition signatures and subsequent tweets were recorded for non-followers of LCV, providing evidence that among that subsample, our manipulation had no effect on these outcomes. There is one significant exception to this finding, however: in Study 1, five users not in the LCV's network *retweeted* either the public tweet or a tweet from one of the followers. In Study 2 this number was 7. With over seven million total followers of the LCV's followers, these magnitudes are minuscule, but they are greater than zero.

Study 2

We implemented a nearly identical research design in Study 2, with some minor improvements to simplify analysis and address the randomization concern described above. To ensure successful randomization in the second stage of the design in Study 2, we used simple random assignment of the tweet encouragement within the survey software (Qualtrics) itself. The three main treatment conditions remained unchanged from Study 1. For those treatments, we used block random assignment by day and number of followers.

We further made two changes to the way the web links (URLs) to the petition worked. First, we were able to incorporate the abbreviated version of the organization's name (LCV) into the URLs themselves in order to boost realism. Second, we passed on the anonymized Twitter IDs of each subject as a query to each URL sent in the DMs so that we could more easily merge individual-level outcomes with treatment assignment.

In Study 2, the subject of the Twitter campaign was more timely: the environmental protection agency's plan to issue regulations mandating reduced carbon emissions from power plants. The public tweet was posted on July 2, 2014, and DMs were sent in 20 batches beginning that day.[10] Between them, the 221 petition signers were followed by 1990 other users, who constitute the subjects of the second-stage experiment in Study 2. Table 3 summarizes the design and basic outcomes of Study 2.

Results: Study 1

Study 1's results challenge the conventional wisdom about Twitter's mobilization capabilities on at least two fronts. Perhaps most surprisingly, not a single subject in the public tweet condition either signed the petition or retweeted the petition link. It is important to reiterate that subjects were exposed to a single public tweet, so this result does not rule out the possibility that a more concerted campaign with multiple tweets might have worked.[11] Further, it is possible that infrequent Twitter users never saw the tweet at all. The ineffectiveness of the public tweet stands in contrast to the strong showing of the direct messages. Without prior research to guide our expectations, we would not have been surprised at either a null finding or a negative "backlash" effect.[12] One alternative interpretation of these results is that DMs are more effective due to repeat exposure: subjects may have responded to the DMs because they had already seen a public tweet featuring the same message, but a tweet alone is not enough to drive outcomes.

A final finding, but one we interpret with caution given the potential imbalance, is that the "organizer" condition caused subjects to send significantly fewer tweets using the randomly assigned tweet encouragement. We detail these findings and discuss the absence of network effects below.

[10] "Follower" condition: "You're one of our most valuable followers! Help fight climate change by signing the petition & tweet to your friends! (URL)"; "organizer" condition: "You're one of our most valuable organizers! Help fight climate change by signing the petition & tweet to your friends! (URL)"

[11] Suppose that the true treatment effect is that a public tweet generates a single click per 10,000 followers exposed. With a sample size of 6687, we would expect to observe zero clicks about 51 % of the time.

[12] These also seemed like plausible results *ex ante*; during a pilot study in which one of the authors sent automated direct messages to a subset of his followers, several recipients warned about "spam" or a possible virus.

Table 3 Study 2: design and outcomes

Treatment group	*N*	Signed (%)	Tweeted (%)
Public tweet	3495	0 (0.0)	0 (0.0)
Organizer DM	2514	107 (4.3)	28 (1.1)
Follower DM	2498	114 (4.6)	36 (1.4)
Total	8507	221 (2.6)	64 (0.8)
Among subjects who signed petition			
Tweet encouragement	111	111 (100.0)	50 (45)
No encouragement	110	110 (100.0)	11 (10)
Total	221	221 (100.0)	61 (27.6)

Main Effects

First, we look at petition signatures as the outcome of interest. As the first two columns of Table 4 show, the "follower" and "organizer" treatments both had positive and significant effects at the $p < .01$ level. The follower DM caused an estimated 3.9-percentage-point increase in the proportion of subjects who signed the petition. The organizer DM caused a 3.3-percentage-point increase in participation; these effects are not significantly different from each other ($p = 0.38$). Thus, the best interpretation of the evidence from Study 1 is that receiving any direct message (after potentially seeing a similar public tweet) caused a 3.6-percentage-point increase in petition signing.

Next, we turn to the tweet outcome. The last three columns of Table 4 show that both DM conditions produced a positive causal effect on tweet activity. The "organizer" message caused fewer tweets than the "follower" message: 1.1 percentage points fewer than the 2.7 percentage point boost generated by the follower message among subjects assigned to the DM conditions ($p = 0.03$). This evidence is suggestive of a priming effect in which the "organizer" identity reduces the future likelihood of tweeting but not the more immediate task of signing an online petition. We return to this apparent finding below.

The second stage of the experiment was designed to identify the causal effect of the tweet button on subsequent tweet activity by the subject's own followers. Table 5 shows that the among those who completed the petition, the effect of being shown a tweet button was large: the treatment caused nearly half of exposed subjects to click and tweet a message about the petition to their followers (the positive constant may seem counterintuitive, but upon investigation we discovered that it reflects users who independently tweeted a link to the petition without using the supplied functionality—either manually or, for example, using a built-in Twitter app in their web browsers. This also illustrates the advantage of an experimental design, which can distinguish between this baseline activity and tweets caused by the manipulation).

The interaction in Model 2 reiterates the previous finding that the organizer message appeared to depress tweet activity. Under an additional assumption[13] that all subjects who signed the petition in one DM condition would have signed the

[13] See Online Appendix 1 for a full discussion of this assumption and its plausibility in this application.

Table 4 Study 1: effects of direct message treatments on participation and tweeting

	Signed		Tweeted	
	(1)	(2)	(3)	(4)
Treatment: follower	0.039***	0.040***	0.027***	0.027***
	(0.005)	(0.005)	(0.004)	(0.004)
Treatment: organizer	0.033***	0.033***	0.016***	0.016***
	(0.005)	(0.005)	(0.003)	(0.003)
Account type: male		−0.004		0.001
		(0.004)		(0.003)
Account type: organization		−0.018***		−0.003
		(0.004)		(0.003)
Account type: unknown		−0.012*		0.004
		(0.007)		(0.007)
Eigenvector centrality		0.002		−0.000
		(0.002)		(0.001)
Number of followers		−0.001		−0.001
		(0.001)		(0.001)
Days on Twitter		−0.001		−0.001
		(0.002)		(0.001)
Days on Twitter missing		0.014		0.010
		(0.014)		(0.015)
Constant	0.000	0.006**	0.000	−0.000
	(0.000)	(0.003)	(0.000)	(0.002)
N	6687	6687	6687	6687
R^2	0.021	0.024	0.014	0.014

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ Robust standard errors in parentheses. Eigenvector centrality, number of followers, and days on Twitter in standard units and centered at zero

Table 5 Study 1: effects of tweet encouragement

	Tweeted	
	(1)	(2)
Shown tweet encouragement	0.454***	0.624***
	(0.086)	(0.104)
Treatment: organizer		0.053
		(0.104)
Encouragement X organizer		−0.384**
		(0.170)
Constant (treatment: no encouragement)	0.186***	0.161**
	(0.051)	(0.067)
N	109	109
R^2	0.214	0.267

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ Robust standard errors in parentheses

petition in the other DM condition, we can interpret the heterogeneous effect in causal terms: being primed as an "organizer" *caused* the encouragement to be less effective to the tune of more than 38 percentage points. In other words, the "follower" prime was more than twice as effective at encouraging future tweets as the "organizer" prime. This is a substantial difference although, again, its interpretation rests crucially on the assumption referenced above.

Network Effects

Recall that the subjects of the second-stage experiment were followers of petition signers. In effect, randomly treating some petition signers with the treat encouragement randomly exposed their followers to additional tweets. Column 1 of Table 6 shows that this manipulation was quite effective—those who followed exposed subjects were 63 percentage points more likely to have seen a retweeted message. However, despite being potentially exposed to retweets, columns two through five show that treated subjects were not significantly more likely to sign or tweet the message themselves. This finding parallels the ineffectiveness of public tweet sent by our partner organization. At least in this experiment, only direct messages had significant effects on participation.

Results: Study 2

The results from Study 2 are broadly in line with those of Study 1.[14] In particular, we replicated the finding that the "organizer" DM condition caused fewer subsequent tweets. The public tweet did not have a significant effect on petition signatures; in this study as well, not a single subject assigned to the public condition completed the petition. As before, additional exposure to direct messages caused a significant number of petition signatures and tweets.

Main Effects

The effect sizes we estimate from Study 2 are somewhat larger than those in Study 1, but they remain substantively comparable.[15] The "follower" and "organizer" messages boosted petition signatures by 4.6 and 4.3 percentage points, respectively (see Table 7). The effects of the direct messages on signing were not significantly different from each other ($p = 0.60$). The direct messages also significantly increased tweet behavior. As in Study 1, priming the "follower" identity was more effective than the "organizer" identity, though the difference is no longer statistically significant.

The tweet link caused a 35.0-percentage-point increase in tweeting behavior in the restricted model shown in Table 8, but, as in Study 1, there were differential effects by DM condition: the button increased tweets by 47.2 percentage points

[14] We did not find any evidence of balance problems as in Study 1.

[15] This could reflect the fact that the campaign was more timely and related to a current political dispute, in addition to the seasonality observed in other types of participation (Rosenstone and Hansen, 1993).

Table 6 Study 1: effects of tweet encouragement on subjects' followers

	Shown tweet	Signed		Tweeted	
	OLS (1)	OLS (2)	IV (3)	OLS (4)	IV (5)
Exposure: followed subject shown tweet encouragement	0.628***	0.005		−0.004	
	(0.029)	(0.006)		(0.010)	
Exposure: followed subject tweeted			0.008		−0.006
			(0.010)		(0.016)
Constant	0.150***	0.010**	0.008	0.012	0.013
	(0.025)	(0.005)	(0.006)	(0.010)	(0.012)
N	1176	1176	1176	1176	1176
R^2	0.374	0.0004	−0.001	0.0003	0.001

Robust standard errors in parentheses. All regressions weighted by inverse probability of exposure

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

among subjects sent the "follower" message and by 23.1 percentage points among subjects sent the "organizer" message. Again, we can interpret this difference causally under the assumption that all petition signers in one DM condition would have signed in the other condition.

Network Effects

In contrast to the null network findings in Study 1, Table 9 presents evidence that signing the petition was strongly influenced by others' tweets. Column 1 shows that the manipulation was effective. Column 2 shows an ITT effect of the tweet encouragement of 2.0 percentage points. Column 3 shows the estimated effect among compliers: subjects who followed others who tweet if and only if they are shown the tweet encouragement were 5.6 percentage points more likely to sign the petition. Considering the relatively low rates of participation generally, these effects are substantively large. Columns 4 and 5 repeat the analyses for the "tweeted" dependent variable: we observe no significant differences by exposure condition at the $p < 0.05$ level. Still, the fact that network effects on petition signatures are larger in magnitude than the main effect of direct contact from LCV itself is broadly consistent with the notion that "networks of recruitment" are important for promoting participation (Verba et al. 1995).

Treatment Effect Heterogeneity by Account Type

Theories of how online mobilization activities affect political behavior are focused on the individual citizen: appeals from peers or groups via social networks may induce citizens to make contributions to public goods. The treatments deployed in our two experiments were developed with individual Twitter users in mind; however, the direct messages were sent to the accounts of organizations as well. All

Table 7 Study 2: effects of direct message treatments on participation and tweeting

	Signed		Tweeted	
	(1)	(2)	(3)	(4)
Treatment: follower	0.046***	0.046***	0.014***	0.014***
	(0.004)	(0.004)	(0.002)	(0.002)
Treatment: organizer	0.043***	0.043***	0.011***	0.011***
	(0.004)	(0.004)	(0.002)	(0.002)
Account type: male		0.005		0.005**
		(0.004)		(0.002)
Account type: organization		−0.019***		−0.003
		(0.004)		(0.002)
Account type: unknown		−0.003		0.002
		(0.008)		(0.004)
Eigenvector centrality		−0.002		−0.001
		(0.001)		(0.001)
Number of followers		0.001		0.003**
		(0.002)		(0.001)
Days on Twitter		−0.002		−0.002*
		(0.002)		(0.001)
Days on Twitter missing		−0.019		0.003
		(0.014)		(0.013)
Constant	0.000	0.003	0.000	−0.001
	(0.000)	(0.003)	(0.000)	(0.001)
N	8507	8507	8507	8507
R^2	0.019	0.022	0.005	0.008

Robust standard errors in parentheses. Eigenvector centrality, number of followers, and days on Twitter in standard units and centered at zero

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

else being equal, we would expect individual users to be more likely to both sign the petition and retweet the petition link.

Twitter does not provide account-type information, so we hand-coded the profiles of all experimental subjects. We coded each account as female, male, an organization, or unknown. We relied on users' profile pictures and descriptions to determine account type. Organizations were easy to identify: they typically use language such as "We are a non-profit dedicated to..." in their description fields. Determining gender could be difficult when the profile pictures were not clearly male or female. When possible, we used cues in the description field such as "Activist, educator, and father of two...". When we could not determine gender or organizational status, we coded a profile as being of unknown type. Two of the authors carried out the coding; on a sample of 200 profiles, our inter-coder reliability was extremely high (Cohen's $\kappa = 0.90$).

Figure 3 presents the results of our heterogeneous effects analyses. The conditional average treatment effects (CATEs) and 95 % confidence intervals are

Table 8 Study 2: effects of tweet button on subsequent tweets

	Tweeted	
	(1)	(2)
Shown tweet encouragement	0.350***	0.472***
	(0.055)	(0.077)
Treatment: organizer		0.037
		(0.059)
Encouragement X organizer		−0.241**
		(0.110)
Constant (treatment: no encouragment)	0.100***	0.083**
	(0.029)	(0.036)
N	221	221
R^2	0.154	0.181

Robust standard errors in parentheses

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 9 Study 2: effects of tweet encouragement on subjects' followers

	Shown tweet	Signed		Tweeted	
	OLS (1)	OLS (2)	IV (3)	OLS (4)	IV (5)
Exposure: followed subject shown tweet encouragement	0.352***	0.020***		0.004	
	(0.031)	(0.006)		(0.003)	
Exposure: followed subject tweeted			0.056***		0.013
			(0.018)		(0.010)
Constant	0.137***	0.010***	0.002	0.004**	0.003
	(0.027)	(0.004)	(0.006)	(0.002)	(0.003)
N	1975	1975	1975	1975	1975
R^2	0.142	0.005	−0.051	0.001	−0.008

Robust standard errors in parentheses. All regressions weighted by inverse probability of exposure

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

shown for all four account types, broken out by dependent variable and study. In Study 1, we observe some treatment effect heterogeneity on the "signed" dependent variable: treatment effects are much smaller for organizations compared to individuals. We observe no such heterogeneity for the "tweeted" dependent variable. The second row presents the estimates for Study 2. We see nearly the identical pattern: on the "signed" dependent variable, organizations have much smaller treatment effects than individuals, but this difference is not apparent for the "tweeted" dependent variable. Interestingly, there is no consistent pattern for the relative size of treatment effects among men and women; the treatments appear to work equally well for both, regardless of dependent variable.

We present these estimates in a regression format in Online Appendix 3, along with heterogeneous effects analyses by subjects' number of followers, number of

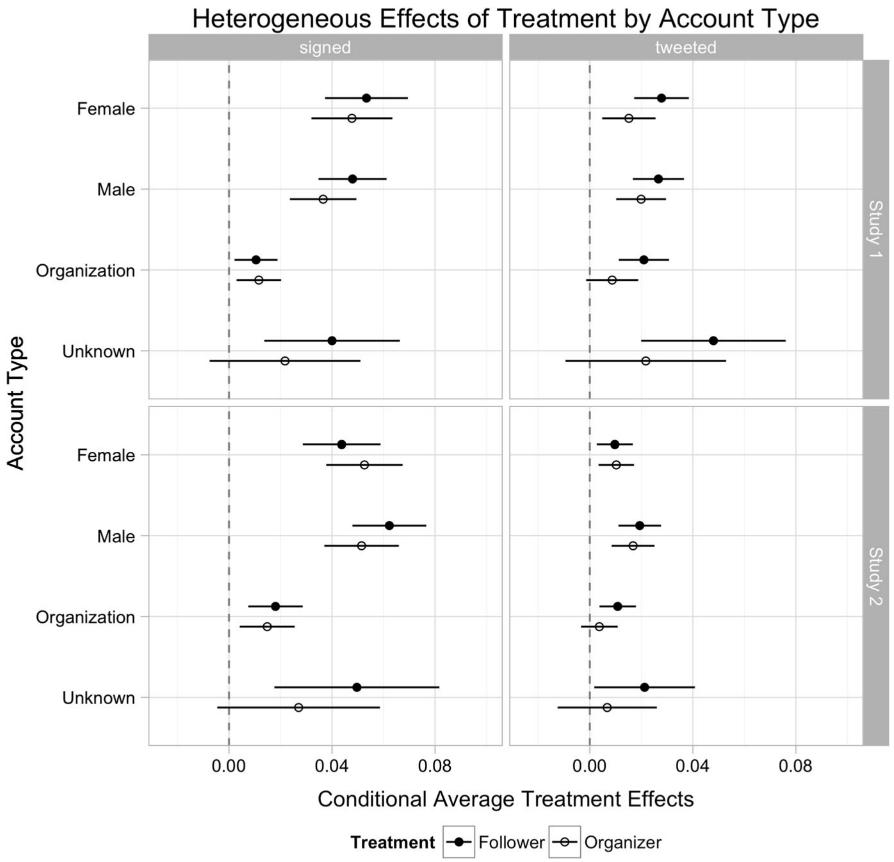

Fig. 3 Entries are conditional differences-in-means with 95 % confidence intervals. In Study 1, the sample was 38.4 % male, 30.4 % female, 24.4 % organizations, and 6.8 % unknown; in Study 2, the sample was 39.3 % male, 32.9 % female, 22.1 % organizations, and 5.7 % unknown

days their Twitter account was active, and eigenvector centrality.[16] These analyses do not uncover a systematic pattern of treatment effect heterogeneity, though it does appear that treatment effects are marginally smaller for more central users. We interpret this finding cautiously, since organizational accounts tend to have higher centrality scores.

Discussion

This study identifies several robust findings about the effectiveness of different types of mobilization appeals on Twitter. First, direct messages are far superior to public tweets in generating supportive behavior in the form of online petition

[16] Online Appendix 2 presents randomization checks using these covariates.

signatures, tweets, or even retweets. In both of our experiments, not a single subject assigned to be exposed only to the public tweet signed or retweeted the petition. We find that DMs produce approximately a 4-percentage-point increase in clicks. If an organization were to send out 250 direct messages (the maximum) per day for 30 days, they could expect to collect $250 \cdot 30 \cdot 0.04 = 300$ signatures over the course of a month. While this is a modest number, it should be weighed alongside the cost per DM, which is effectively zero.

Our results speak to the literature on social media and collective action. We find some support for the "reconceptualized" collective action theory of Bimber et al. (2005) when comparing the magnitudes of our effects. DMs of both types cause an increase in petition signatures greater than the effect on overall tweets to the petition link. However, compared to the effect of randomly assigning subjects who had already completed the petition to see the tweet button (35 to 45 percentage points), we see that the act of sending out a public tweet is arguably much easier to induce than a petition requiring some time or effort to complete. While this is also consistent with some notions of "slacktivism," we find that overall hypothesis difficult to square with the null effect of the public tweet on retweets in both studies. Overall, the results seem most consistent with the traditional perspective elaborated in the Civic Voluntarism Model, in which network effects are most effective.

Designing experiments on a social network like Twitter is difficult for a number of reasons. Most apparent is the fact that public tweets are potentially visible to anyone, which makes identifying their effects impossible without imposing additional assumptions about over-time persistence and anticipation. Another issue inherent to Twitter is a lack of individual-level exposure measures: even if there existed a reliable indicator for whether a tweet was potentially visible to a given user (perhaps because a mobile or desktop app was active at the time), it would still greatly overstate whether it was actually seen and retained. This means that practically speaking, the only available estimand will be ITT.

One exception is cases in which the treatment is an encouragement to tweet and compliance can easily be measured. Our second-stage experiments, in which subjects were followers of the tweeters, have precisely this design. Its advantage is in distinguishing between the effects of homophily—similar users may already follow each other, be interested in similar issues, and retweet each other's updates—from the effects of contagion (McPherson et al. 2001; Fowler et al. 2011). In Study 2 but not Study 1, we find that inducing tweet behavior within the network causes a significant and substantively large number of additional petition signatures. This is a potentially important finding for organizations seeking to launch "viral" campaigns: public tweets may be more effective when sent by followers of an organization than by the organization itself. Alternatively, users may require repeated exposure to public tweets in order for them to be effective. However, since this finding did not replicate across both studies, it should be interpreted with caution.

Regardless, messages to followers are effective when they take the form of private DMs. While this may seem counterintuitive given Twitter's public network structure, one possible explanation is that individualized contact gives these messages the same essential properties as email (most users in fact receive an email notification when sent a DM). Email can be ineffective for certain purposes, such as mobilizing voter turnout (Nickerson 2007). But if a message is perceived as *solicited* contact from a trusted source, we hypothesize that it can be effective. This is consistent with existing research on emailed recruitment messages for web-based surveys, which emphasizes the torrent of unsolicited email and spam that users face daily. As one study points out (Porter and Whitcomb 2003), despite the relative ease with which spammers can mimic other senders, "it is still difficult to change the credibility of the message itself" (p. 587). In this case, the credibility lies in the fact that the recipient has already chosen to follow updates from the originator of the message (albeit over another medium).

Of course, we cannot rule out the possibility that these effects depend on multiple exposures to the same message (via an initial public tweet and subsequent DM). Future designs might vary exposure to public tweets over time in order to better address this question. It is also possible that public tweets work differently than posts on other social networks, such as Facebook. While we cannot test this possibility here, it seems plausible that the sheer number of tweets posted in real time diminish the effectiveness of any single post, while Facebook's algorithms keep the amount of social content to a manageable level, thus boosting the impact of any individual item. Experimental research has found strong effects of get-out-the-vote posts on Facebook, for example (Teresi and Michelson 2014).

Our final result is the differential effect of the "follower" and "organizer" identity primes. Existing research in social psychology has established that priming individual traits can have subsequent, unconscious effects on behavior. This can also extend to the ways in which people perceive themselves: When priming specific values (e.g., caring about the environment), subjects will tend to adjust their choices and behavior accordingly—but only if those values are central to their self-concept (Verplanken and Holland 2002). The implications for this study are straightforward. If members of an engaged network of environmental activists view themselves as organizers, priming this identity could bring forward other relevant considerations, such as the commitment it entails.

The differential effects of the "organizer" versus "follower" messages on the probability of tweeting may shed light on two theoretical questions. The first is, How does the authenticity of a message change its effectiveness? Twitter users may find the "organizer" label disingenuous, because in reality, they just subscribe to the advocacy group's Twitter feed. The second question is, Are messages that prime the costs of collective action less effective? Perversely, encouraging subjects to help overcome free-rider problems with costly actions may primarily serve to reiterate that grassroots organizing is indeed personally taxing.

This is a surprising possibility given the traditional expectation that the Internet has the ability to promote collective action by reducing transaction costs overall (Farrell 2012). Twitter clearly possesses the properties—such as speed, reach, and versatility—necessary for this to be the case. Despite these low structural costs, organizations nevertheless compete for individuals' limited attention online. Even small changes in perceived costs can reduce the probability of collective action.

Acknowledgments The authors would like to thank the League of Conservation Voters, whose support and eagerness to include new experimental designs in its programs made this study possible. We also thank the Analyst Institute for its initial support of this project. The authors' names are listed in alphabetical order. We particularly thank Amit Mistry, Vanessa Kritzer, Kristin Brown, and Mo Maraqa at LCV for their crucial roles in the planning and implementation of the two experiments. We are further grateful to Andy Przybylski and Neelanjan Sircar for extremely helpful discussions and feedback during the early stages of this project. We also thank Kevin Collins, Lindsay Dolan, Ben Farrer, Don Green, Lucas Leemann, three anonymous reviewers, and the editors of *Political Behavior* for their helpful comments and advice. An earlier version of this paper was presented at the 2014 Annual Meeting of the American Political Science Association in Washington, D.C., where we benefited from a lively discussion with participants and comments by Jaime Settle. This study would not have been possible without the use of Columbia University's Hotfoot High Performance Computing (HPC) Cluster for hosting the scraping programs.

References

Aaker, J., & Akutsu, S. (2009). *Why do people give?*. The role of identity in giving: Stanford University Graduate School of Business Research Paper.

Barr, D., & Drury, J. (2009). Activist identity as a motivational resource: Dynamics of (dis)empowerment at the G8 direct actions, Gleneagles, 2005. *Social Movement Studies*, *8*(3), 243–260.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Bennett, W. L., & Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, *15*(5), 739–768.

Bimber, B., Flanagin, A. J., & Stohl, C. (2005). Reconceptualizing collective action in the contemporary media environment. *Communication Theory*, *15*(4), 365–388.

Bowers, J., Fredrickson, M. M., & Panagopoulos, C. (2013). Reasoning about interference between units: A general framework. *Political Analysis*, *21*(1), 97–124.

Bryan, CJ., Walton, GM., Rogers, T., Dweck, CS. (2011). Motivating voter turnout by invoking the sel. In Proceedings of the National Academy of Sciences (vol 108(31), pp. 12,653–12,656).

Converse, P. E. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), *Ideology and Discontent*. Ann Arbor: University of Michigan Press.

Coppock, A. (2014). Information spillovers: Another look at experimental estimates of legislator responsiveness. *Journal of Experimental Political Science*, *1*(02), 159–169.

Drury, J., Cocking, C., Beale, J., Hanson, C., & Rapley, F. (2005). The phenomenology of empowerment in collective action. *British Journal of Social Psychology*, *44*(3), 309–328.

Farrell, H. (2012). The consequences of the internet for politics. *Annual Review of Political Science*, *15*, 35–52.

Fowler, J. H., Heaney, M. T., Nickerson, D. W., Padgett, J. F., & Sinclair, B. (2011). Causality in political networks. *American Politics Research*, *39*(2), 437–480.

Gaby, S., & Caren, N. (2012). Occupy online: How cute old men and malcolm X recruited 400,000 US users to OWS on Facebook. *Social Movement Studies*, *11*, 367–374.

Gerber, A. S., & Green, D. P. (2000). The effects of personal canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review*, *94*, 653–663.

Gerber, A. S., & Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.

Gladwell, M. (2010). Small change: Why the revolution will not be tweeted. The New Yorker.

Gong, S., Zhang, J., Zhao, P., Jiang, X. (2014). Tweets and sales. Retrived from SSRN Working Paper.

Karpf, D. (2010). Online political mobilization from the advocacy group's perspective: Looking beyond clicktivism. *Policy & Internet*, *2*(4), 7–41.

Kobayashi, T., Ichifuji, Y. (2014). Tweets that matter: Evidence from a randomized field experiment in Japan. In *Annual Meeting of the Southern Political Science Association*.

Krueger, B. S. (2006). A comparison of conventional and internet political mobilization. *American Politics Research*, *34*(6), 759–776.

Lupia, A., & Sin, G. (2003). Which public goods are endangered?: How evolving communication technologies affect the logic of collective action. *Public Choice*, *117*(3–4), 315–331.
Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133.
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444.
Morozov, E. (2009). Iran: Downside to the 'Twitter revolution'. Dissent.
Nickerson, D. W. (2007). The ineffectiveness of e-vites to democracy: Field experiments testing the role of e-mail on voter turnout. *Social Science Computer Review*, *25*(4), 494–503.
Obar, J., Zube, P., & Lampe, C. (2012). Advocacy 2.0: An analysis of how advocacy groups in the United States perceive and use social media as tools for facilitating civic engagement and collective action. *Journal of Information Policy*, *2*, 1–25.
Olson, M. (1965). *The logic of collective action*. Cambridge: Harvard University Press.
Porter, S. R., & Whitcomb, M. E. (2003). The impact of contact type on web survey response rates. *Public Opinion Quarterly*, *63*, 579–588.
Rosenstone, S., & Hansen, J. M. (1993). *Mobilization, Participation and Democracy in America*. New York: MacMillan Publishing.
Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations*. New York: Penguin Group.
Shulman, S. W. (2009). The case against mass e-mails: Perverse incentives and low quality public participation in US federal rulemaking. *Policy & Internet*, *1*(1), 23–53.
Sinclair, B., McConnell, M., & Green, D. P. (2012). Detecting spillover effects: Design and analysis of multilevel experiments. *American Journal of Political Science*, *56*(4), 1055–1069.
Taylor, S.J., Muchnik, L., Aral, S. (2014). Identity and opinion: A randomized experiment. Retrieved from SSRN.
Teresi, H., Michelson, M. (2014). Wired to mobilize: The effect of social networking messages on voter turnout. *The Social Science Journal*. doi:10.1016/j.soscij.2014.09.004.
Verba, S., Schlozman, K. L., & Brady, H. E. (1995). *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge: Harvard University Press.
Verplanken, B., & Holland, R. W. (2002). Motivated decision making: Effects of activation and self-centrality of values on choices and behavior. *Journal of Personality and Social Psychology*, *82*(3), 434.