Supporting Information

Testing Social Science Network Theories with Online Network Data: An Evaluation of External Validity

James Bisbee and Jennifer M. Larson

Contents

1 Survey Design

1.1 Survey Questions

Many questions are accompanied by a scale, a field in which to rank names, or a diagram to populate.

1. Do you use an online social media product? (i.e., Facebook, Twitter, Instagram, Google +, etc.)

   ELICITATION

2. Please provide the name (first name only) of a contact with whom you interact less than once per month on average [online/offline]. This person should be an acquaintance and not a friend.

3. Please provide the name (first name only) of a contact with whom you interact once per month or more [online/ offline]. This person should be an acquaintance and not a friend.

4. Please provide the name (first name only) of a contact with whom you interact less than once per week on average [online/ offline]. This person should be a friend and not an acquaintance.

5. Please provide the name (first name only) of a contact with whom you interact once per week or more [online/ offline]. This person should be a friend and not an acquaintance.

   The following questions describe hypothetical scenarios involving the contacts you rely on for financial support. Please read each question carefully and answer as truthfully as possible.

   INFORMATION

6. Imagine you are looking for a new job. Please rank your social ties in order of who you would ask for leads.

   TRUST

7. Imagine you are starting a project and in need of funding. Please rank your social ties in order of who you believe would be most interested in contributing.

8. Imagine you are starting a project and in need of funding. Please rank your social ties in order of who you believe would most effectively garner the support from others.

   The following questions relate to a hypothetical scenario involving $100 that is given with no strings attached. This money is given freely by an anonymous third party and the recipient can do with it what (s)he wants. Please read the following questions carefully and answer as truthfully as possible.

   DONATE

9. Imagine you are the recipient of the $100. Please indicate how you would divide the $100 between yourself and your five contacts.

TRUST

10. Now imagine one of your contacts is the recipient of the $100 and is asked the same question as above. Please rank your contacts in order of who you would want to play the game such that you would receive the most money.

EGALITARIANISM

11. Now imagine one of your contacts is the recipient of the $100 and is asked the same question as above. Please rank your contacts in order of who you would want to receive the money such that every member of the group would receive some amount of money.

The following questions are designed to examine the how similar you are to your social ties and how well you know them. The first four questions ask you to identify the social tie you feel you are most similar to along four different characteristics (political views, religious views, educational attainment, and life skills). The following two questions ask you to identify which social tie you would want to play a game with in which you earn money for each of their responses you correctly guess.

HOMOPHILY

12. Which of your social ties do you feel you are most similar to in terms of political ideology?

13. Which of your social ties do you feel you are most similar to in terms of religious beliefs?

14. Which of your social ties do you feel you are most similar to in terms of educational attainment?

15. Which of your social ties do you feel you are most similar to in terms of life skills?

16. Imagine you are asked to play a game in which you can stand to win money for each question you answer correctly. The questions are about what answer your social tie gives to a question on their political opinions. If you correctly guess the response of your social tie, you earn money. For example, the question might be "Does Person Y support legislation that would allow companies to drill for oil in Alaska?" If you correctly guess Person Y's response, you would earn $10.

    Among your social ties, who would you choose to play this game with in order to maximize your earnings? (In other words, whose political views do you feel that you know the most about?)

17. Imagine you are asked to play a game in which you can stand to win money for each question you answer correctly. The questions are about what answer your social tie gives to a question on their religious views. If you correctly guess the response of your social tie, you earn money. For example, the question might be "Does Person Y believe in reincarnation?" If you correctly guess Person Y's response, you would earn $10.

    Among your social ties, who would you choose to play this game with in order to maximize your earnings? (In other words, whose religious views do you feel that you know the most about?)

The following questions examine how you would relate to different people in different contexts. The first four questions require you to sort your contacts in different scenarios. The next pair of questions asks you to place your contacts around a dinner table to achieve different results. The final question asks you to explicitly sort your contacts by the strength of the connection to them.

INTIMACY

18. Think about your contacts in the context of a professional crisis (i.e., the loss of a job). Please categorize your contacts in terms of who you would notify in order of priority and those who you would not ever discuss your crisis with. NOTE: think of your response in terms of one-on-one communication and aggregated information (i.e., general Facebook posts or public Tweets).

19. Think about your contacts in the context of a personal crisis (i.e., the death of a loved one). Please categorize your contacts in terms of who you would notify in order of priority and those who you would not ever discuss your crisis with. NOTE: think of your response in terms of one-on-one communication and not social media or aggregated information (i.e., general Facebook posts or public Tweets).

20. Think about your contacts in the context of a professional success (i.e., a promotion or new job). Please categorize your contacts in terms of who you would notify in order of priority and those who you would never notify. NOTE: think of your response in terms of one-on-one communication and not social media or aggregated information (i.e., general Facebook posts or public Tweets).

21. Think about your contacts in the context of a personal success (i.e., a new relationship). Please categorize your contacts in terms of who you would notify in order of priority and those who you would never notify. NOTE: think of your response in terms of one-on-one communication and not social media or aggregated information (i.e., general Facebook posts or public Tweets).

The following two questions as you to place your contacts at a dinner table. For the purpose of these questions, please consider vertical links (i.e., seated across from an individual) as those with the strongest interactions. In addition, consider horizontal links (i.e., seated next to an individual) as those with the second strongest interactions. Finally, consider diagonals as the weakest interactions. The diagram below summarizes these assumptions with the thickness of the arrow representing the strength of the interaction.

COMMONALITY

22. Imagine you are planning for a dinner party such as at a wedding. The seating chart for yourself and your five contacts is given below. Please place yourself and your five social relations around the table such that you maximize the interactions between people who have the least in common. Please only place one person at a seat.

23. Imagine you are planning for a dinner party such as at a wedding. The seating chart for yourself and your five contacts is given below. Please place yourself and your five social

relations around the table such that you maximize the interactions between people who have the most in common. Please only place one person at a seat.

STRENGTH

24. Please rank your contacts in order of tie strength from strongest (1st) to weakest (5th).

INTERACTION

25. In a perfect world, how often would you like to interact with your social contacts?

PROTEST - PEACEFUL

The following questions relate to a hypothetical scenario in which [RANDOMLY ASSIGNED TIE] is protesting a controversial bill currently up for a vote in the legislature that you oppose. [RANDOMLY ASSIGNED TIE] calls you to ask if you'll join in the protest. The protest is peaceful and legal.

26. Would you join the protest?

27. Which, if any, of your other social ties would you inform about the protest in an attempt to get them to participate?

28. Which, if any, of your other social ties would prompt you to join the protest if you knew they were also participating?

PROTEST - VIOLENT

The following questions relate to a hypothetical scenario in which [RANDOMLY ASSIGNED TIE] is protesting a controversial bill currently up for a vote in the legislature that you oppose. [RANDOMLY ASSIGNED TIE] calls you to ask if you'll join in the protest. The protest is not sanctioned by the local authorities and there have been reports of violent clashes between some protesters and the police.

29. Would you join the protest?

30. Which, if any, of your other social ties would you inform about the protest in an attempt to get them to participate?

31. Which, if any, of your other social ties would prompt you to join the protest if you knew they were also participating?

EQUIVALENCY TESTS

32. Please select which of the following social networking sites, if any, you use regularly (at least a few times per year) and drag them into the box provided. Within this box, please rank those that you use by order of frequency with 1 being the most frequent. If you do not use a particular platform, do not add it to the box.

33. Please select which of the following email platforms, if any, you use regularly (at least a few times per year) and drag them into the box provided. Within this box, please rank those that you use by order of frequency with 1 being the most frequent. If you do not use a particular platform, do not add it to the box.

34. Other types of online social networks can develop around other types of online platforms. For example, raiding parties in Massively Multiplayer Online Games (MMOs) can be considered a social network. Another example may be the communities that develop around online forums, such as the Turker boards. Do you engage in these types of online social networks? If so, please name them in the fields below (up to 5). If you do not participate in one of the communities more than a few times per year, leave all boxes blank.

35. The follow question asks you to indicate the frequency of interaction you share with each of your social ties on [MOST FREQUENTLY USED SOCIAL NETWORKING SITE]. These interactions can be comments, likes, messages, shares, endorsements, tweets, retweets, or any other method of interaction allowed by [MOST FREQUENTLY USED SOCIAL NETWORKING SITE].

36. The follow question asks you to indicate the frequency of interaction you share with each of your social ties on [MOST FREQUENTLY USED EMAIL PLATFORM]. These interactions can be comments, likes, messages, shares, endorsements, tweets, retweets, or any other method of interaction allowed by [MOST FREQUENTLY USED EMAIL PLATFORM].

37. The follow question asks you to indicate the frequency of interaction you share with each of your social ties on [MOST FREQUENTLY USED ONLINE COMMUNITY]. These interactions can be comments, likes, messages, shares, endorsements, tweets, retweets, or any other method of interaction allowed by [MOST FREQUENTLY USED ONLINE COMMUNITY].

DEMOGRAPHICS

38. Please answer the following demographic questions. Please select your highest educational attainment category.

39. What is your age?

40. In which industry are you employed? (U.S. Census)

41. Please indicate your occupation:

42. What is your race?

43. In which country do you reside?

44. What is your family structure? (U.S. Census)

45. What is your gender?

46. What is your combined annual household income?

47. What is your current status?

48. Please indicate the extent to which you agree with the following questions. I prefer one-on-one conversations to group activities.

49. I often prefer to express myself in writing.

50. I enjoy solitude.

51. I seem to care about wealth, fame, and status less than my peers.

52. People tell me that I'm a good listener.

53. I'm not a big risk-taker.

54. I like to celebrate birthdays on a small scale, with only one or two close friends or family members.

55. People describe me as "soft-spoken" or "mellow."

56. I often let calls go through to voice-mail.

1.2 Variable Glossary

With 17 separate measures of tie strength, we provide a simple glossary of terms. We divide our measures into broad categories corresponding to different dimensions of theoretical interest. Note, however, that we do not assume that these dimensions are mutually exclusive. Below, we discuss the implications of correlated measures on family wise error rates when testing whether each measure differs significantly between the online and offline elicitation treatments.

Furthermore, we are sensitive to the challenges associated with relative explanatory power when treating these measures as right hand side variables. Below, we discuss our motivation for using structural equation modeling (SEM) to guard against spurious Type I errors as described by Westfall and Yarkoni (2016). Table 1 summarizes our measures.

Table 1: Glossary of Terms

Dimension	*Label*	*Question*
Homophily	Political Homophily	Which of your social ties do you feel you are most similar to in terms of political ideology?
	Religious Homophily	Which of your social ties do you feel you are most similar to in terms of religious beliefs?
	Educational Homophily	Which of your social ties do you feel you are most similar to in terms of education attainment?
	Class Homophily	Which of your social ties do you feel you are most similar to in terms of life skills?
	Least in Common	(Seating Chart) Please place yourself and your five social relations around the table such that you maximize the interactions between people who have the least in common.
	Most in Common	(Seating Chart) Please place yourself and your five social relations around the table such that you maximize the interactions between people who have the most in common.
Reciprocity	Contribute	Imagine you are starting a project and in need of funding. Please rank your social ties in order of who you believe would be most interested in contributing.

Table 1: Glossary of Terms (continued)

Dimension	*Label*	*Question*
	Garner Contributions	Imagine you are starting a project and in need of funding. Please rank your social ties in order of who you believe would most effectively garner the support of others.
	Personal Gain	Now imagine one of your contacts is the recipient of the $100 and is asked the same question as above (Donation Game). Please rank your contacts in order of who you would want to play the game such that you would receive the most money.
	Group Gain	Now imagine one of your contacts is the recipient of the $100 and is asked the same question as above (Donation Game). Please rank your contacts in order of who you would want to play the game such that every member of the group would receive some amount of money.
Intimacy	Personal Crisis	Think about your contacts in the context of a personal crisis (i.e., the death of a loved one). Please categorize your contacts in terms of who you would notify in order of priority and those you would not ever discuss your crisis with.
	Personal Success	Think about your contacts in the context of a personal success (i.e., a new relationship). Please categorize your contacts in terms of who you would notify in order of priority and those you would never notify.
	Preferred Interaction	In a perfect world, how often would you like to spend time with your social contacts?
	Tie Strength	Please rank your contacts in order of tie strength from strongest (1st) to weakest (5th).
Professional	Job Search	Imagine you are looking for a new job. Please rank your social ties in order of who you would ask for leads.
	Professional Crisis	Think about your contacts in the context of a professional crisis (i.e., the loss of a job). Please categorize your contacts in terms of who you would notify in order of priority and those who you would not ever discuss your crisis with.
	Professional Success	Think about your contacts in the context of a professional success (i.e., a promotion or new job). Please categorize your contacts in terms of who you would notify in order of priority and those who you would not ever discuss your crisis with.

2 Specification

2.1 Characteristics Comparisons

Our baseline specification for the difference in network measures is given in equation 1:

$$M_{i,t} = \alpha + \rho_t D_i + \epsilon_{i,t} \tag{1}$$

where $M_{i,j}$ is the measure of interest for subject i relative to tie t, D_i is a dummy indicator taking on the value of 1 if the subject was randomly assigned to the online condition and 0 otherwise, and $\epsilon_{i,t}$ is an error term. In this experiment, ρ_t measures the difference in the tie measure between the online and offline conditions.

2.2 Relationships Comparisons

We also compare the relationship between tie strength measures and outcomes across the online and offline elicitation conditions. To do this, we first calculate the relationship presented in equation 2 separately for both the online and offline conditions.

$$Y_{i,t} = \alpha + \beta_t M_{i,t} + \eta_{i,t} \tag{2}$$

As in equation 1, $M_{i,t}$ is the tie strength measure for subject i relative to her connection to tie t. We capture the relationship between this measure and an outcome of interest ($Y_{i,t}$, the donation amount in the dictator game) with β. We calculate a vector of $\hat{\beta}$'s for both the online and offline conditions and then test the difference of these point estimates using a t-test $\left(\frac{\hat{\beta}_{on} - \hat{\beta}_{off}}{\sqrt{se^2_{on} + se^2_{off}}}\right)$.

2.3 Model Comparisons

Finally, we compare online and offline relationships using cross-validation to characterize the comparability of models used to analyze social network questions. Specifically, we fit a variety of models to the online data and use the models to predict outcomes in the offline context. We calculate the root mean squared error (RMSE), the mean absolute error (MAE), and the Bayesian Information Criterion (BIC) associated with the online/offline division of the data.

We then compare these measures of model fit to those produced by a similar cross-validation technique that randomly divides the dataset in half instead of by online/offline source. We repeat this technique 1,000 times for each behavioral outcome (donation and protest), each tie, and each model and save the RMSE, MAE, and BIC from each simulation.

All three measures are widely used in empirical studies to compare different models. In our application, we hold the models constant and instead compare the online/offline dataset division to 1,000 simulated results generated by dividing the data in half a random. To formalize our comparison, we report the percent of the 1,000 simulated results which yield higher RMSE, MAE, and BIC values than the online/offline division. Although not a proper confidence threshold, we interpret percentages analogous to p-values from a conventional hypothesis test, highlighting those cases in which the random divisions yield superior measures of model fit more in more than 950 simulations.

2.4 Consistency and Null Results

Since we randomly assign subjects to elicitation condition, our causal interpretation of the results does not require conditioning on observables. However, given the substantive importance of our null findings, we want to ensure that insignificant differences between elicitation conditions are not the product of noisy estimates. Following Angrist and Pischke (2008), we improve the consistency of our estimates via controlling for covariates.

An ideal approach would be to design our experiment such that we improve consistency via block randomization. Given the empirical context, we deemed this approach too costly as it would involve first collecting a sample of online respondents in one period and recording their demographic characteristics, and then asking them to return for a second survey at a different period for block randomization. Instead, we employ a variety of methods to improve consistency.

The simplest approach is to include demographic characteristics as controls in a multiple regression framework (Equation 3). Doing so yields adjusted averages of our potential outcomes via Frisch-Waugh-Lovell which reduces the variation in $\hat{\rho}$.

$$M_{i,j} = \alpha + \rho D_i + \beta \mathbf{X}_i + \epsilon_i \tag{3}$$

We also employ other techniques that pursue the same improvements in consistency without sacrificing as many degrees of freedom. First, we use inverse propensity score weighting (IPW) in which we use pre-treatment covariates as predictors in assignment to treatment and then calculate the inverse probability of treatment via

$$w_i = \frac{D_i}{\pi_i} + \frac{1 - D_i}{1 - \pi_i} \tag{4}$$

where π_i is the propensity of assignment to treatment calculated using Bayesian Additive Regression Trees (BART)(Hill, 2012).[1]

Second, we use Coarsened Exact Matching (CEM) to replicate the benefits associated with block randomization *post hoc*. Following Iacus, King and Porro (2011), we first coarsen our observed pre-treatment covariates into bins and then use these bins to identify stratification cells. Within each cell, we weight the control observations such that their weighted total equals the number of treated observations in that cell, and then estimate the ATT as a simple difference in these weighted potential outcomes.

2.5 Incremental Validity

In our main results, we apply the various specifications described above to a simple regression of donation decisions on a single measure of tie strength. In the analyses below, we dive deeper into the data and explore whether different measures of tie strength are more prognostic of behaviors of interest. Traditional methods of adjudicating between the relative strength of competing covariates use "horse race" regressions in which competing covariates are added and removed from regressions and the change in relevant coefficients is interpreted. These interpretations typically follow the

[1] We use BART instead of a more conventional logit due to (1) the ability of BART to identify the most prognostic covariates without using unnecessary degrees of freedom and (2) the freedom from parametric assumptions required when using a logit.

argument that, if a covariate remains significant even after controlling for other independent variables in a multiple regression, it must make a unique contribution to the outcome. Or conversely, variables whose significance and coefficient decline with the inclusion of a new control must not be prognostic explanatory variables.

Westfall and Yarkoni (2016) find that horse race methods of adjudicating between competing independent variables lead to explosive Type I errors, particularly when the measured covariates are considered proxies for latent characteristics measured with some error. In these cases, Type I error rates exceed the stated α and are increasing in (1) the "unreliability" of the proxy measure, (2) the indirect effect between the latent characteristic and the outcome, and the sample size. "[M]easurement unreliability makes it easier for the regression model to confuse the direct and indirect paths... The larger the influence of the confounding covariate, the more variance can be misattributed to the predictor of interest." (Westfall and Yarkoni, 2016, p. 13). These authors demonstrate the robustness of structural equation models (SEM) to inference issues associated with incremental validity.

Our 17 measures of tie "strength" are designed to capture different dimensions of this theoretically-vital concept. However, we readily admit that our measures are only proxies for the underlying dimensions of interest and that their reliability may vary. As such, we adopt SEM in our multiple regression analyses that compare the predictive power of all 17 dimensions at once. We also include more traditional horse race regressions for comparison.

2.6 Family-Wise Error Rates and Type I Error

We also confront a different source of Type I error in our analysis. We follow convention and concern ourselves with a confidence threshold of 95%, corresponding to a 5% chance of falsely concluding that there is a significant relationship in the data. However, with multiple comparisons, the true probability of committing a Type I error increases dramatically. Specifically, with 17 different dimensions tested, we would expect roughly 1 estimate to be statistically significant at the 95% confidence level (0.05 * 17). The family-wise error rate (FWER) is simply the combined probability of committing a Type I error when testing multiple hypotheses and can be expressed as:

$$FWER = 1 - (1 - \alpha)^m \tag{5}$$

where m is the number of tests. With a two-tailed test and 17 separate hypotheses, our FWER is $1 - (1 - 0.025)^{17} = 0.35$, much higher than we feel comfortable basing conclusions on. As such, we adjust our threshold for each individual hypothesis using a variety of methods.

The simplest method is a Bonferroni correction that accounts for the increased risk of committing Type I errors (Dunn, 1961). Specifically, the correction establishes a p-value threshold of $p_h \leqslant \frac{\alpha}{m}$ where subscript h indexes each hypothesis. In our context, the adjusted threshold for statistical significance is therefore $p_h \leqslant \frac{0.025}{17} = 0.0015$. Figure 1 depicts the effect of elicitation on measures of tie strength using both conventional confidence intervals (depicted in light gray) and Bonferonni corrected intervals (white). These results corroborate our analysis conducted in the paper.

However, the Bonferroni Adjustment assumes that outcomes are independent, resulting in an overly conservative test for situations in which the independence assumption is violated. Given that we designed our survey to capture different but related dimensions of social ties, we do not believe in the independence assumption on theoretical grounds, an intuition confirmed in our analysis of

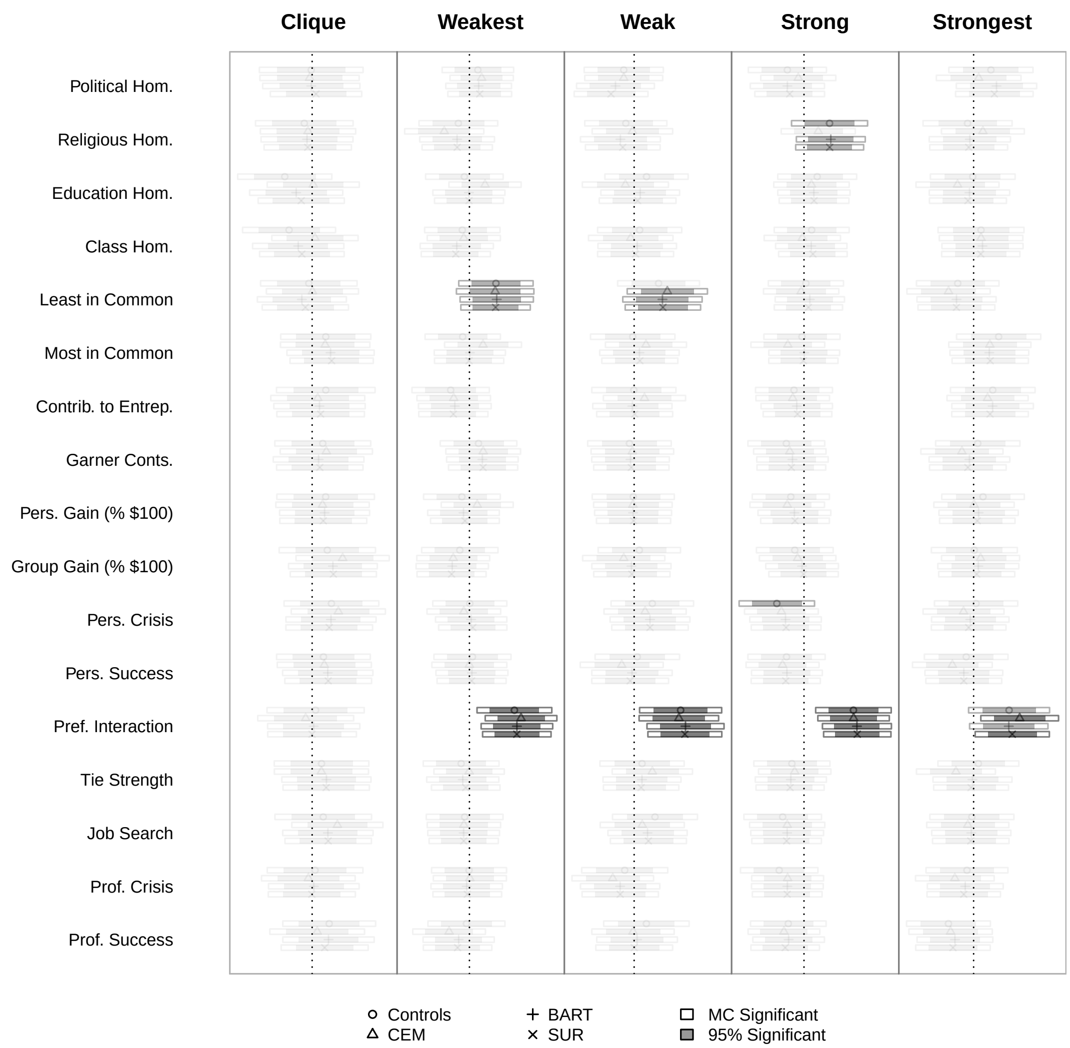

Figure 1: Raw measures of tie strength dimensions are regressed on the online elicitation treatment via four different specifications: multiple regression analysis, coarsened exact matching, propensity score matching via BART, and seemingly unrelated regressions. Dimensions are presented on the y-axis. The figure charts the coefficient estimate, the 95% confidence interval (darker bars), and the Bonferroni adjusted confidence intervals (white bars). Light shades of gray indicate non-significant treatment effects, darker indicate significance at the unadjusted 95% level, and black indicates significance after applying the Bonferroni adjustment.

the correlation presented in Figure 3 of the main text.

Following Westfall and Young (1993), we adopt a step-down approach in which we bootstrap our sample and compare the significance of enforced null results to the significance of our models using the full data. This approach is less conservative than the Bonferroni adjustment as it appropriately accounts for correlated outcomes. In practice, both the Bonferroni and step-down resampling approaches yield very similar results and the choice of adjustment is inconsequential to our conclusions. As such, our paper uses these adjustments interchangeably.

An alternative approach to accounting for correlated outcomes is to use Seemingly Unrelated

Regressions (SUR) (Zellner, 1962). SUR addresses two concerns associated with our analysis. First, in accounting for correlated outcomes across multiple tests, the method speaks to the Type I error concerns discussed in the preceding section. Second, the method further improves on the consistency of estimates by characterizing correlated disturbance terms across tests. Our conclusions are robust to the use of SUR specifications.

3 Equivalency Tests

Our results rest on the assumption that respondents understood the elicitation as intended. There may be concern that our subjects did not notice or appreciate the randomized assignment to the online or offline elicitation frameworks and only provided the first five names that came to mind. If this was the case, our null results only reflect the failure of our elicitation framework and cannot speak to more substantive conclusions regarding the mapping between online and offline social network data. In the section that follows, we present equivalency tests that suggest this concern is unfounded.

3.1 Elicitation Differences

In the article text, we show that none of the measures of tie strength dimensions differ between the online and offline contexts, with the exception of the preferred level of interaction. We posit that this supports the internal validity of our elicitation framework and speaks to the substantive topic of online/offline parity. Importantly, the significant results for the preferred level of interaction persist even under the most conservative tests that account for the Family-Wise Error Rate (shown in Figure 1 above).

Given the significant relationship between treatment and preferred level of interaction, we can say with confidence that our elicitation framework captures a meaningful difference in the preferred level of interaction between online and offline networks. More forcefully, we appeal to the identifying assumption of random assignment to elicitation and conclude that, at minimum, subjects who saw the word "online" want to spend more time with their social contacts than those who saw the word "offline". Assuming that subjects read and understood the elicitation framework and honestly reported real social ties, we can further conclude that subjects want to see their online social ties more than their offline.

3.2 Equivalency Tests

However, we are not yet able to determine whether this difference in preferred interaction is the product of online and offline ties being different individuals or simply a priming effect from elicitation. It is possible that subjects associate the word "online" with suboptimal interaction levels and therefore report a desire for more frequent interaction. In this scenario, it may be that the subjects still failed to understand the elicitation instructions and the significant difference we document is simply a priming effect.

To test this, we asked subjects to indicate how much time they currently interact with their ties on a variety of online social networks. This question allows us to directly compare the amount of time respondents currently spend interacting with their ties online *across elicitation frameworks.* If our survey failed to elicit genuine ties, the results for online and offline ties would look equivalent. As illustrated in Figure 2, this is not the case.

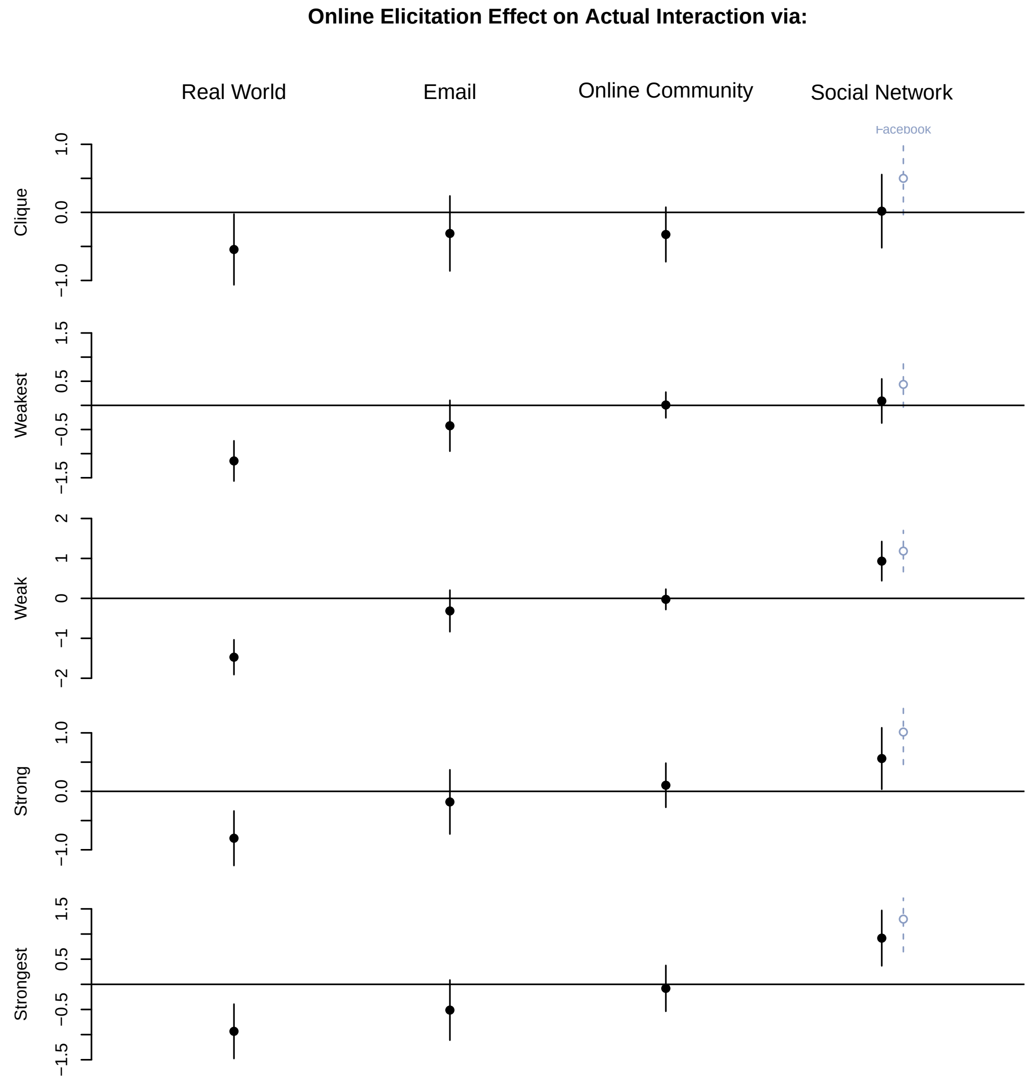

Figure 2: Differences in measures of self-reported interaction in the real world, via email, in online communities (MMOs, virtual communities, etc.), and social networks (Facebook, Twitter, Instagram) between treated and control groups.

In particular we note significant differences in the frequency of interaction for the strongest and the weak ties between online and offline elicitation. Specifically, subjects interact more frequently with their weak, strong, and strongest online ties than their offline via online social networks (these include Facebook (80%), Twitter (8%), Instagram (5%), and LinkedIn (2%)). We don't see a significant difference between the online and offline elicitation frameworks for the weakest and clique elicitations although restricting attention only to Facebook (the most popular online social network in our sample) pushes all estimates well beyond or up to the 95% level of confidence.

In addition to the systematically different levels of online interaction, we also measured the average real world interaction in both elicitation frameworks. Here, we ask subjects to estimate the frequency with which they interact with their ties in offline contexts, such as at the office, at each other's home's, at bars or schools, or at church. The difference between the elicitation frameworks is even more striking in this context, with subjects reporting significant lower frequencies of interaction with their social ties from the online elicitation relative to the offline. In this test, only the clique tie is not significant at conventional levels of significance. All other estimates are significant at even the most conservative adjustment for multiple comparisons.

This interpretation carries substantive meaning for how we understand the parity we document between online and offline social networks. Specifically, it suggests that the names provided by subjects exposed to the online elicitation are not the same as those they would have provided in the offline elicitation. Yet, while the identities of these ties differ, we see no difference in the measures of tie strength we collect, the relationship between these measures and social behaviors, or the residuals generated by regression models relative to random cross-validation tests.

3.3 Platforms

A final concern with our survey results is that the "online" environment we capture is not related to those typically used by researchers. Our primary contribution to the literature is evidence of empirical parity between online and offline sources of social networks. However, if our online elicitation captures social networks on private platforms such as email, our contribution is diluted. Reassuringly, we find strong support for the online social networks we survey as mapping onto existing sources of social network data.

The equivalency tests used above also provide information on the types of online social media subjects use. It may be that subjects consider platforms such as email as online social networks. While not substantively uninteresting, email is less relevant for our methodological contribution as it is rarely used to gather data on social networks due to privacy restrictions. Furthermore, the motivating literature, particularly that which deals with protests, typically focuses on platforms such as Twitter and public Facebook groups.

While we do document a substantial use of email platforms by our subjects, we also note robust use of other platforms such as Facebook and Twitter. In particular, only one respondent reported using only email. All other subjects use a combination of email and some other online social network more conducive to data gathering for researchers interested in measuring social networks. Interestingly, we do not document a systematic difference in frequency of interaction via email platforms over different elicitation frameworks

In the interest of speaking to other subcultures that exhibit online social networks, we also ask users if they participate in communities that develop around other types of online activities. Examples include online computer games, fan clubs for celebrities, product forums, or online help pages. Given the plethora of these types of pages, we ask respondents to write in up to five examples, ordered by frequency of use.

We note a large percentage of subjects listing this third type of online community (43%). We speculate that this reflects a characteristic specific to our sampling pool of Turkers and constitutes a limit on our ability to extrapolate to larger populations. Specifically, Turkers have their own online community built around their profile on Amazon Mechanical Turk and are known to participate regularly in Turker forums. On these forums, users discuss particularly profitable jobs, share information on certain tasks, and rank the quality of the posters. Turkers are also ranked by those who post the jobs they work on and each Turker profile is ranked according to their performance on jobs. By sampling only from the Turker population above a certain rank, we are likely oversampling from individuals predisposed to participate in this third type of online social network. Indeed, of the 43% reporting participating in this third type of online community, the majority of respondents listed one of the Turker forums.

We do not document a statistically significant difference between the frequency of interaction via these types of online communities across our elicitation frameworks. This suggests that the social ties provided by subjects in response to our elicitation framework are not the same as those populating these online forums.

Nevertheless, we argue that our fundamental conclusions are not threatened by our sample since the ties we elicit and the online platforms via which our subjects interact with these ties extend beyond the Turker social network. We believe that Turkers are an informative sample on which we can make inferences about broader social networks and we refer to a range of existing literature that has replicated the results of behavioral studies using the Turker population. We leave an appraisal of our external validity to the reader and limit our causal claims to the subpopulation of Turkers from which we sample.

3.4 Access

Although social network platforms like Facebook and Twitter offer unparalleled control over sampling network data, many sources restrict access to these novel new data. For example, full Facebook data is made available only to data scientists working at the company. Twitter APIs are capped either by rate or by number of tweets. In the case of the Twitter streaming API, the sampling methodology for how tweets corresponding to a certain keyword is unknown.

While these obstacles reduce the appeal of online social network data somewhat, similar challenges occur with the most forms of personal or private data. Firms withhold financial information, Census data is limited at finer levels of aggregation, and researchers face stringent requirements from institutional review boards (IRBs). Overcoming the barriers to access associated with online social network data is a challenge but one that we believe is an order of magnitude less costly than the logistics and expenses associated with gathering commensurate network data via traditional methods.

Nevertheless, it is important to remain cautious about the trajectory of research using these new data sources. In particular, variation in ease of access that is endogenous to the research topic is its own form of external validity limitation. In the context of this paper, we appreciate these concerns but view them as second-order. If online social networks are fundamentally dissimilar to offline social networks, variation in access is moot. We hope this paper helps clear away this first-order concern and opens the door to more rigorous analysis of whether and how empirical results covary with access.

4 Ecological Validity

In our main results, we analyze the parity of relationships between dimensions of tie strength and a hypothetical outcome: the share of $100 that the subject would donate to each of her ties. We find that the relationships are similar in the online and offline elicitation contexts, leading us to conclude that it is not just the measures of social ties but also how they interact with behaviors of interest that persists in both online and offline data.

However, given the hypothetical nature of this behavior, one may question the ecological validity of our findings. There may be concern that the lack of actual costs and benefits associated with our outcomes could elicit meaningless responses, resulting in random associations that are equally random in the online and offline elicitations. In the following section we present evidence in favor of the ecological validity of our outcome measures by (1) comparing the behavior of our subjects to existing research using real money and (2) documenting systematic patterns of prognostic dimensions of tie strength that are inconsistent with purely random responses.

4.1 Existing Experiments

Our donation measure is fundamentally similar to a canonical laboratory experiment called the Dictator Game. In this game, subjects are tasked with dividing a finite sum of real money between themselves and another subject. The original version (Kahneman, Knetsch and Thaler, 1986) of this game found evidence of altruistic behavior, contrary to the theoretical assumptions of self-interest that underpinned the dominant game-theoretic models of behavior. Since then, hundreds of studies have been run that adjust various parameters to examine how norms and self-interest combine. A meta analysis of this literature was conducted by (Engel, 2011) and we compare our donation results to this paper to validate our approach.

Our donation game, despite being purely hypothetical, yields measures in line with laboratory experiments in which subjects are incentivized with real money. Specifically, the share of subjects who keep all of the hypothetical money in our context (24%) is almost identical to the original dictator game (26%). Relative to the meta analysis, our estimate is within 4 percentage points (28.3%) of the meta proportion of subjects who act purely out of self-interest. The share of our subjects who donate the entire amount (5.1%) is also commensurate to laboratory experimental results (ranging between 5% and 10%) (see Frohlich, Oppenheimer and Moore (2001) and Eckel and Grossman (1996)) and almost identical to the meta estimate of 5.04%. The total amount donated by our subjects (46%) is toward the upper tail of the meta distribution but this is likely due to our framework using the real friends and acquaintances of our subjects as opposed to the strangers used in most laboratory applications.

We can use a subset of the meta analysis that focuses on contexts more comparable to our experiment. Specifically, we look at average donation amounts for non-students (39.8%) which is the more appropriate comparison for our Turker population. Similarly, the magnitude of our outlier declines when comparing our estimate to dictator games with multiple recipients (meta average of 39.2%).

It is worth mentioning that other researchers have supplemented real-money laboratory experiments with hypothetical questionnaires. (Brañas-Garza, 2006) tests the similarity in distributions between laboratory and hypothetical experiments, finding that the results are comparable. The author uses this finding to justify merging the datasets and his ensuing contribution is based on the aggregated data that combines both real money and hypothetical measures.

We do not attempt to contribute to behavioral economic research using our hypothetical results. Nevertheless, the parity of our measures with existing laboratory research suggests that the donation game we use to compare online and offline relationships is, at minimum, consistent with existing research that uses real money. Insofar as the reader finds our analysis of tie strength dimensions compelling, we hope that our contribution can spur future work that uses more ecologically valid measures of human behavior.

4.2 Non-Random Responses

Although the analysis summarized above is consistent with the ecological validity of our donation game, it only compares altruistic behavior in our hypothetical setting to similar behavior in lab experiments using real money. The hypothetical nature of our survey may lead to spurious null comparisons between relationships if there is no systematic relationship between different dimensions of tie strength and donation behavior. In other words, while subjects might feel morally obligated to donate a certain percentage consistent with other experiments using real money, they might choose who to donate to at random, leading to our false conclusion that therefore online and offline social networks are interchangeable.

There are a few different ways of confirming that respondent allocation decisions are non-random despite the hypothetical nature of our survey. The simplest is to check whether the elicited ties are good predictors for donation amounts. If subjects are taking the question seriously, we should expect to see less donated to weaker ties relative to strong. As illustrated in Figure 3, this is indeed the case with subjects donating significantly larger shares of the $100 to their strong and strongest elicited ties.

We can also explore the extent to which different dimensions of tie strength predict donation decisions. Our main analysis presented a battery of bivariate regressions exploring the relationship between tie strength dimension and donation decisions by elicited tie. Here, we extend this analysis to look for the most prognostic dimension even after controlling for other dimensions. As discussed above, we adopt a Structural Equation Model (SEM) approach to guard against spurious associations, the results of which are summarized in Figure 4.

We find compelling evidence in favor of the dimensions of preferred interaction, personal gain, political homophily, and trust as proxied with who the subject would share a personal crisis with. These dimensions are intuitive, particularly the personal gain measure which captures the expected reciprocity of the beneficiary. Furthermore, the use of SEM allows us to conclude that these significant findings are incrementally valid: they contribute uniquely to the decision to donate to each tie type, even after controlling for competing dimensions of the relationship.

The consistent support for preferred interaction is an interesting result in its own right, suggesting that this type of survey item is better able to plumb the emotional connection between social ties. We leave a deeper exploration of factor loadings to future research, only using these results to demonstrate that subjects appear to take the donation game seriously. The systematic relationships between dimensions and donation decisions are inconsistent with the concern that subjects were responding at random after allocating a particular amount to themselves.

5 Protest Analysis

Our survey presented subjects with a hypothetical vignette of a protest and asked them whether they would join were they invited by a randomly chosen tie. The question of how social networks influence protest decisions is a particularly salient topic in studies of collective behavior. In particular, determining the type of the social connection most prognostic of participation is an open question in the literature.

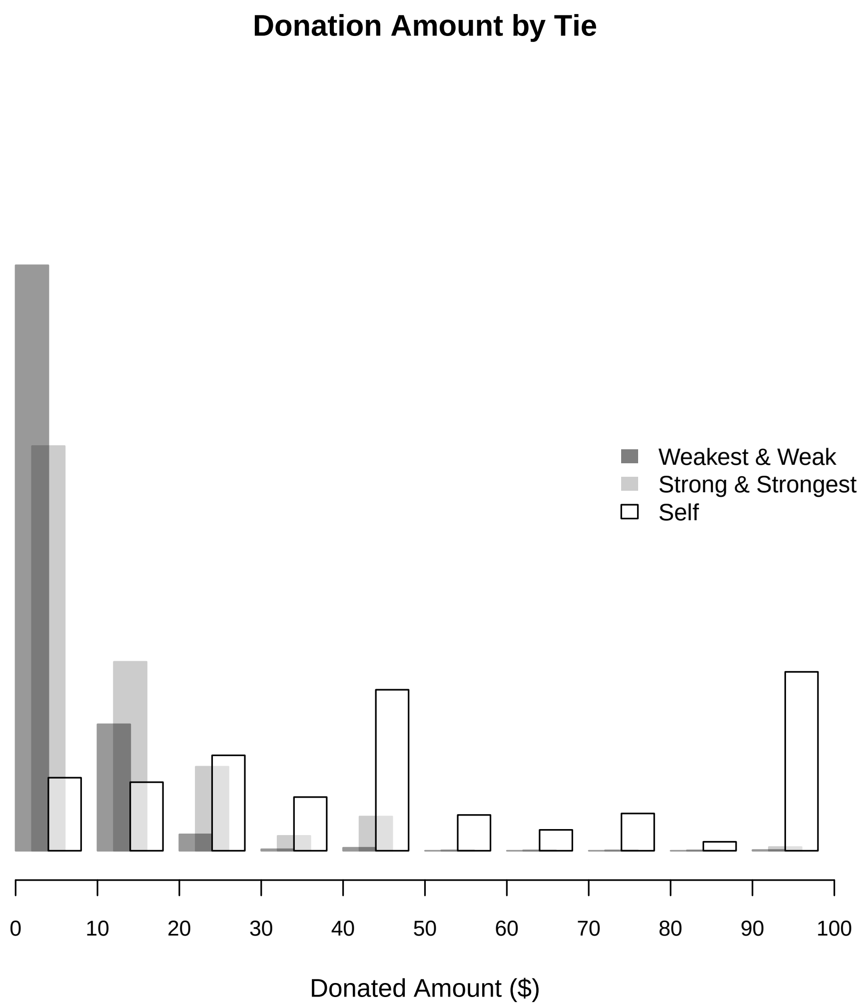

Figure 3: Histograms of donation amounts by elicited tie type and kept for self. T-statistics for difference-in-means between strongest and weakest = 12.7, strong and weak = 5.3, and strongest and strong versus weakest and weak = 13.1.

As with the donation game, we recognize that the purely hypothetical nature of our survey limits our ability to draw substantive conclusions about costly expressions of collective action. Protests involve a variety of costs, including opportunity costs, social sanctioning by peers who disagree with the protest, and in some cases the threat or realization of physical harm. However, we believe that the causal identification afforded by our survey's random assignment of the inviting tie is worth a detailed analysis.

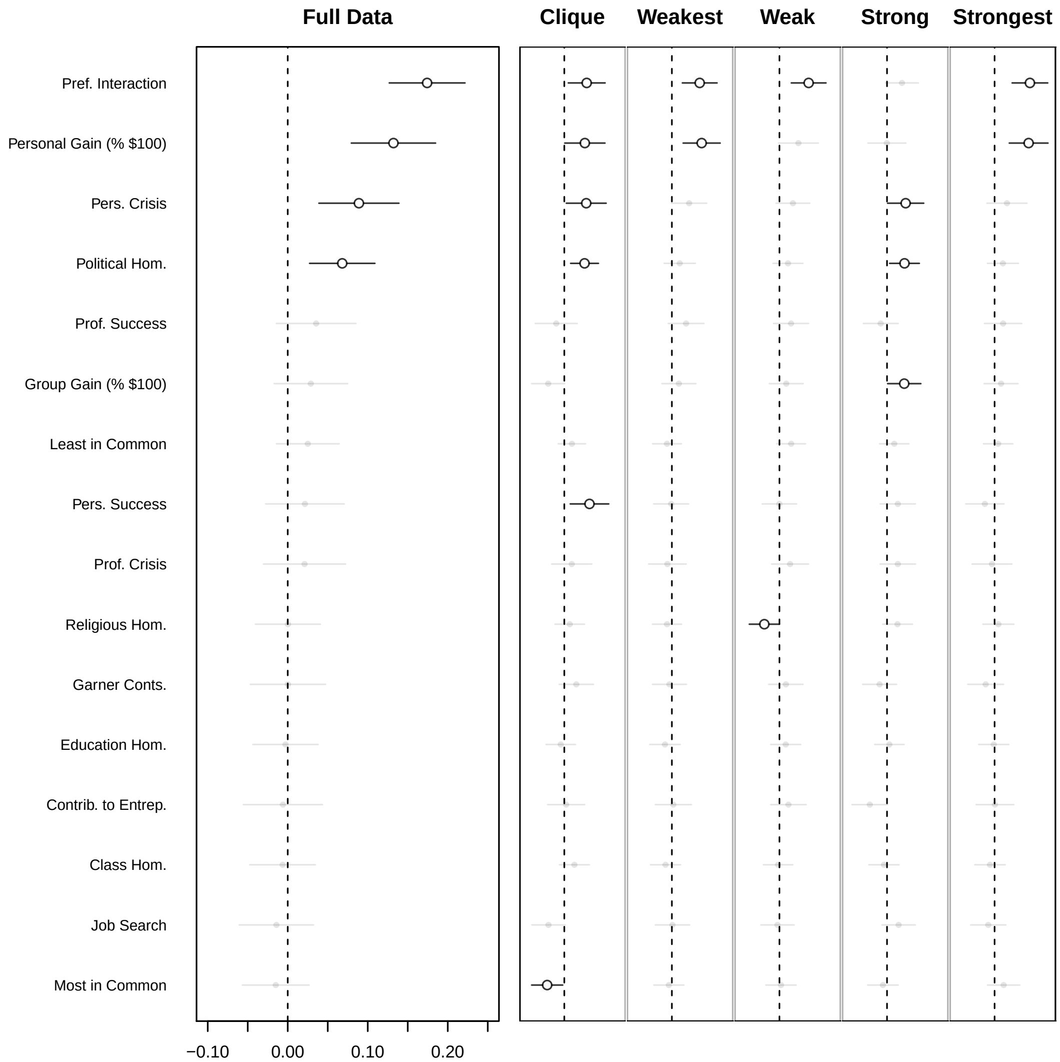

Figure 4: Structural Equation Models (SEM) of the determinants of donating a portion of \$100 to each tie. The main panel presents the results estimated using the full data while the small panels on the right summarize the results for each elicited tie type in isolation.

5.1 Correlations

Before looking at the causal results, we start with the hypothetical questions to look for evidence of non-random responses. In addition to being presented with an invitation to join a protest from a randomly selected tie, subjects were also asked to name one of their ties whose invitation would prompt them to join. While these relationships are not causal, we review the results to confirm that systematic patterns persist. Table 2 explores how the political homophily covariate changes as we increase the number of competing variables in the model. We find consistent significant results for political homophily across all specifications. Substantively, we find that a standard deviation increase in the political homophily between the subject and her tie corresponds to a 9% increase in the probability that the respondent will join the protest. This result remains significant even after

Table 2: Respondent's Decision to Join a Protest Regressed on Tie Strength Dimensions

	(1) Biv.	(2) Cont. 1	(3) Cont. 2	(4) CEM 1	(5) CEM 2	(6) Full	(7) SEM
Political Hom.	.14***	.14***	.14***	.15***	.12***	.10***	.09***
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Peaceful Protest			-.02	-.02	-.02	-.03	-.03
			(.04)	(.04)	(.04)	(.03)	(.03)
Online			-.01	-.02	-.01	-.00	.00
			(.04)	(.04)	(.04)	(.04)	(.03)
Pref. Interaction					.04	.04	.03
					(.02)	(.02)	(.02)
Job Search					.03	.02	.01
					(.02)	(.02)	(.02)
Contrib. to Entrep.					.00	-.00	.01
					(.03)	(.02)	(.02)
Garner Conts.					.05**	.05**	.04**
					(.02)	(.02)	(.02)
Personal Gain (% $100)					.03	.02	.03
					(.03)	(.03)	(.02)
Group Gain (% $100)					-.01	-.01	-.01
					(.03)	(.02)	(.02)
Prof. Crisis					-.02	-.03	-.02
					(.03)	(.03)	(.02)
Prof. Success					.03	.04	.02
					(.03)	(.02)	(.02)
Pers. Crisis					.01	.02	.01
					(.03)	(.02)	(.02)
Pers. Success					.04*	.04*	.05**
					(.02)	(.02)	(.02)
Religious Hom.					.01	.01	.01
					(.02)	(.02)	(.02)
Education Hom.					.02	.03*	.02
					(.02)	(.02)	(.02)
Class Hom.					-.01	-.01	-.00
					(.02)	(.02)	(.02)
Least in Common					.02	.02	.01
					(.02)	(.02)	(.02)
Most in Common					.00	-.00	-.00
					(.02)	(.02)	(.02)
Strong Tie						.21***	.20***
						(.06)	(.05)
Strongest Tie						.40***	.40***
						(.05)	(.05)
Weak Tie						-.02	-.04
						(.05)	(.05)
Weakest Tie						-.05	-.06
						(.05)	(.05)
N	680	668	668	628	581	581	615

Notes: Heteroskedastic-robust standard errors presented in parentheses. Column (1) presents the bivariate linear probability regression of joining a protest on political homophily. Columns (2) through (3) add demographic controls for the respondent. Column (4) uses coarsened exact matching to pair treated to control respondents based on demographic covariates. Columns (5) and (6) add other measures of tie strength. Column (7) re-estimates the model via SEM. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

including a dummy for the identity of the inviting tie, along with all competing measures of tie strength, using a structural equation model. This estimate is roughly double the size of all other competing dimensions with the exception of the identity dummy.

We document similar albeit weaker relationships in Table 3. Again, political homophily is an intuitive predictor of both the subject's decision to join a protest as well as her choice of which tie to inform about a protest should she choose to attend. The systematic patterns exhibited in Tables 2 and 3 reassure us that subjects took the hypothetical vignette of joining a protest seriously. While we do not claim that our results can speak to the decisions of subjects confronting an actual survey, we at least are confident in the internal validity of our estimates.

5.2 Causal Effects

However, the analysis above uses open-ended questions in which the subject is asked to identify which of their ties would prompt them to (1) join a protest and, conditional on their joining (2) who else they would invite. To explore the causal relationship between tie strength and behavior, our survey introduced the protest section with a vignette that was followed by a randomly selected tie who hypothetically invites the join to join them in the protest. This question preceded the open-ended questions summarized above, guarding against priming.

Figure 5 summarizes our findings. The left panel presents the naive bivariate regressions of each dimension of the randomly assigned inviting ties on the subject's probability of joining the protest, controlling only for the nature of the protest itself (peaceful vs. violent). As illustrated, almost all dimensions of tie strength are positive and significant with the exception of the least in common measure which is intuitively negative.

However, the right panel finds only two significant dimensions after controlling for all dimensions together using SEM. Interestingly, we see that the dimension associated with looking for employment is negatively related to the decision to join a protest. This may reflect the belief that one's employment prospects would be hurt if their participation in a protest was known.

Also interesting is the insignificant estimate for political homophily, despite its prevalence in the correlations analyzed above. This contrast may reflect priming or experimental desireability biases in which the subject selects ties most politically similar when asked open-ended questions about who to involve in a protest. Yet when the ties are randomly assigned, reciprocity is the characteristic most prognostic of the decision to join.

As above, we restrain ourselves from making broader claims about these findings when applied to more ecologically valid datasets. Nevertheless, we believe our sacrifice of ecological validity in return for causally-identified evidence of tie dimension heterogeneity is valuable. In addition, by being able to randomly assign the identity of an inviting individual, we are able to highlight an area of substantive interest in the protest literature itself. Future research that exploits a natural experiment to get similar random variation in the identities of inviters to real-world protests should pay careful attention to the measures of tie strength they use as explanatory variables, in particular measures of political homophily and reciprocity.

6 Cliques

We have concluded that there is no systematic evidence of differences between online and offline data in terms of (1) the measures of tie strength we elicit, (2) the relationships between these measures and behaviors of interest, and (3) the fit of models tested on online data used to predict offline outcomes relative to random divisions of the data.

Table 3: Respondent's Decision to Inform Regressed on Tie Strength Dimensions

	(1) Biv.	(2) Cont. 1	(3) Cont. 2	(4) CEM 1	(5) CEM 2	(6) Full	(7) SEM
Political Hom.	.09***	.09***	.09***	.08***	.04**	.05**	.05***
	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)
Peaceful Protest			.01	-.02	-.02	-.02	.02
			(.05)	(.05)	(.04)	(.04)	(.04)
Online			-.00	.01	.03	.03	.02
			(.04)	(.04)	(.04)	(.04)	(.03)
Pref. Interaction					.05**	.05**	.06***
					(.02)	(.02)	(.02)
Job Search					.05**	.05**	.04*
					(.02)	(.02)	(.02)
Contrib. to Entrep.					.03	.03	.03
					(.03)	(.02)	(.02)
Garner Conts.					.05**	.04*	.02
					(.02)	(.02)	(.02)
Personal Gain (% $100)					.03	.03	.04
					(.03)	(.03)	(.02)
Group Gain (% $100)					.01	.01	.02
					(.02)	(.02)	(.02)
Prof. Crisis					.04	.03	.03
					(.03)	(.03)	(.02)
Prof. Success					.05**	.05*	.05**
					(.02)	(.02)	(.02)
Pers. Crisis					.04	.04	.03
					(.03)	(.03)	(.02)
Pers. Success					.01	.01	.00
					(.03)	(.03)	(.02)
Religious Hom.					.02	.01	.01
					(.02)	(.02)	(.02)
Education Hom.					.01	.02	.02
					(.02)	(.02)	(.02)
Class Hom.					-.01	-.01	-.02
					(.02)	(.02)	(.02)
Least in Common					-.00	-.01	-.00
					(.02)	(.02)	(.02)
Most in Common					.00	.01	.01
					(.02)	(.02)	(.02)
Strong Tie						.08	.07
						(.06)	(.05)
Strongest Tie						.31***	.31***
						(.05)	(.05)
Weak Tie						.01	.01
						(.06)	(.05)
Weakest Tie						-.12**	-.13**
						(.05)	(.05)
N	684	652	652	624	569	569	597

Notes: Heteroskedastic-robust standard errors presented in parentheses. Column (1) presents the bivariate linear probability regression of joining a protest on political homophily. Columns (2) through (3) add demographic controls for the respondent. Column (4) uses coarsened exact matching to pair treated to control respondents based on demographic covariates. Columns (5) and (6) add other measures of tie strength. Column (7) re-estimates the model via SEM. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

However, we have noted that these conclusions are less robust when we restrict our analysis to the elicited clique tie. Cliques generally had more dimensions enter significant when looking at the

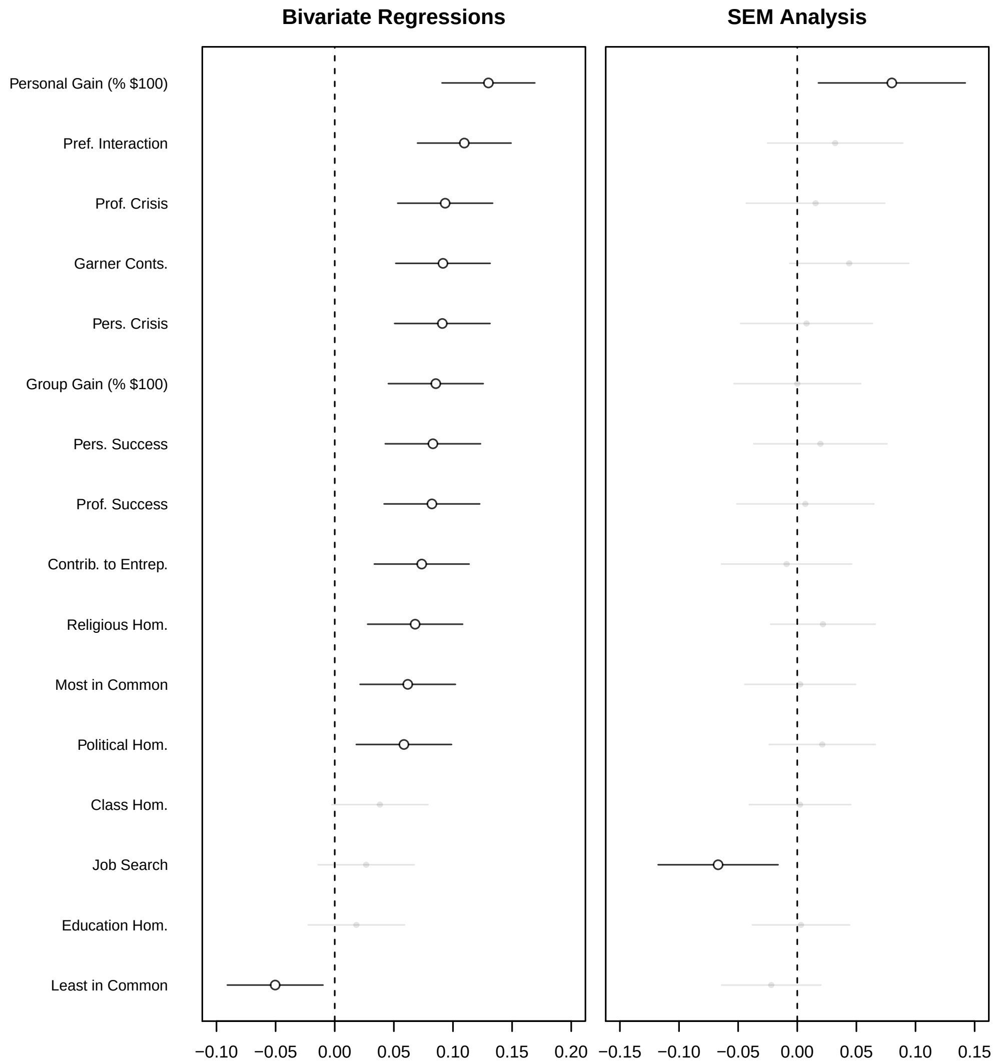

Figure 5: Relationship between tie strength dimensions of the randomly assigned tie who invites the subject to a protest and the probability of joining. The left panel presents the results of simple bivariate comparisons taking each dimension in isolation. The right panel uses a Structural Equation Model to account for relative influence on the decision.

relative prognostic strength of all 17 measures on donation and protest outcomes. Cliques were the one tie type for whom the robust findings for preferred level of interaction was not significant and for whom the elicitation checks were not robust. And cliques were the one tie type for whom the model fit comparisons suggested systematic improvements using random divisions of the data.

We believe these findings are consistent with friends-of-friends being more diffuse in online data, yielding greater variation in the strength of the social connection between our subjects and the cliques they chose. This greater variation (1) leaves greater explanatory power for the dimensions we measure to predict subject behavior toward this tie and (2) makes the online/offline model fits inferior to random divisions of the data.

References

Angrist, Joshua D and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Brañas-Garza, Pablo. 2006. "Poverty in dictator games: Awakening solidarity." *Journal of Economic Behavior & Organization* 60(3):306–320.

Dunn, Olive Jean. 1961. "Multiple comparisons among means." *Journal of the American Statistical Association* 56(293):52–64.

Eckel, Catherine C and Philip J Grossman. 1996. "Altruism in anonymous dictator games." *Games and economic behavior* 16:181.

Engel, Christoph. 2011. "Dictator games: A meta study." *Experimental Economics* 14(4):583–610.

Frohlich, Norman, Joe Oppenheimer and J Bernard Moore. 2001. "Some doubts about measuring self-interest using dictator experiments: The costs of anonymity." *Journal of Economic Behavior & Organization* 46(3):271–290.

Hill, Jennifer L. 2012. "Bayesian nonparametric modeling for causal inference." *Journal of Computational and Graphical Statistics* .

Iacus, Stefano M, Gary King and Giuseppe Porro. 2011. "Causal inference without balance checking: Coarsened exact matching." *Political analysis* p. mpr013.

Kahneman, Daniel, Jack L Knetsch and Richard H Thaler. 1986. "Fairness and the assumptions of economics." *Journal of business* pp. S285–S300.

Westfall, Jacob and Tal Yarkoni. 2016. "Statistically controlling for confounding constructs is harder than you think." *PloS one* 11(3):e0152719.

Westfall, Peter H and S Stanley Young. 1993. *Resampling-based Multiple Testing: Examples and Methods for p-Value Adjustment*. Vol. 279 John Wiley & Sons: New York.

Zellner, Arnold. 1962. "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias." *Journal of the American statistical Association* 57(298):348–368.