

IDENTIFYING PEER EFFECTS IN ONLINE COMMUNICATION

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF COMMUNICATION

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Dean Eckles

June 2012

© 2012 by Dean Griffin Eckles. All Rights Reserved.
Re-distributed by Stanford University under license with the author.

This dissertation is online at: <http://purl.stanford.edu/sf348ds5644>

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Clifford Nass, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Jeremy Bailenson

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Art Owen

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Byron Reeves

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

David Rogosa

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumpert, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

Understanding how the behavior of individuals is affected by the behavior of their peers is of central importance for social and behavioral science. Most of the credible evidence about peer effects currently comes from agent-based simulations and small experiments in constructed social environments, as until recently investigators have been largely unable to quantitatively study these processes at a large scale and *in situ*. Despite massive new data sources, credible identification and estimation of peer effects has remained a largely unsolved methodological problem. Peer effects can produce clustering of a behavior of interest in social networks, but so can other processes, including homophily and the consequences of prior peer influence. This presents major challenges for identifying, bounding, or even determining the sign of peer effects in observational studies. While experiments are an appealing alternative, it is not often possible to conduct field experiments that identify peer effects by directly manipulating peer behavior. Sometimes investigators can manipulate particular mechanisms of peer effects or manipulate network formation processes, but not only are these situations rare, but, these experiments do not always identify effects of primary interest. Thus, better understanding of the promise and limitations of quasi-experimental and observational methods are needed.

This dissertation evaluates and advances the methods for identifying peer effects,

with particular focus on applications to studying online communication behavior. We review of theories of peer effects multiple fields, highlighting their common interest in learning about individuals' exposure–adoption functions. We systematically review the assumptions required for identification of peer effects. We then use reanalysis of a large experiment that blocked a critical mechanism of peer effects as a constructed observational study; that is, we use methods for observational data to reanalyze the experiment while replacing the experimental control group with a nonexperimental control group. Treating the experimental results as a gold standard, we find that propensity score methods can eliminate most of the confounding bias, at least when good measures of closely related prior behaviors are available for use in the propensity score model. These results provide evidence that observational methods may sometimes provide informative estimates of peer effects.

We also present the use of peer encouragement designs as a promising identification strategy for peer effects. In peer encouragement designs, experimenters manipulate assignment of an individuals' peers to an encouragement to the focal behavior. This peer encouragement is a shock to peer behavior that in turn affects ego behavior; we argue such encouragements often make valid and informative instrumental variables for peer effects. We illustrate these methods with a large encouragement design on Facebook in which individuals were randomly assigned to be prompted to express gratitude on Thanksgiving Day 2010. We find evidence of positive peer effects in expressions of gratitude via Facebook. A purely observational analysis produces smaller estimates of peer effects than the instrumental variables analysis using the peer encouragement design, likely reflecting differences in the population of individuals these analyses average over. We argue that encouragement designs are an underused research design for studying peer effects.

Acknowledgements

My doctoral studies were advised by Clifford Nass, who I thank for his always insightful advice on research, teaching, and life. The breadth and depth of Cliff’s knowledge in social science and methodology has always been impressive. This — and his unwavering enthusiasm for scientific discovery — made him an invaluable resource throughout my doctoral studies.

I also benefited from interaction with colleagues in the Communication between Humans and Interactive Media (CHIME) Lab. I thank Erina DuBois, who deserves much credit for making CHIME an organized, productive, and fun environment for research. I was very fortunate that Maurits Kaptein became a repeat visitor to CHIME, as our close collaboration was both directly productive and more generally important to my intellectual development. I thank Stanford Media X and Philips Research for creating this opportunity. Throughout my doctoral studies I enjoyed the fellowship of my cohort — Key Lee, Solomon Messing, Ethan Plaut, James Scarborough, Kathryn Segovia, and Sean Westwood.

I arrived at the topic of this dissertation in part through encounters with product design decision-making at Facebook and interaction with a diverse set of colleagues there. At Facebook, I have been lucky to collaborate with and receive critical assistance in this work from Eytan Bakshy, Jonathan Chang, Danny Ferrante, Adam

Kramer, Thomas Lento, Cameron Marlow, Michael Nowak, Flavio Oliviera, Itamar Rosenn, Keith Schacht, Gabriel Trionfi, and Jeffrey Wieland. The link sharing experiment of Chapter 4 was conducted by Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic, who each provided guidance in its reuse as a constructed observational study. Eytan in particular was generous in helping me understand details of this experiment and providing comments and suggestions that substantially improved this work. The Thanksgiving Day experiment of Chapter 5 was implemented by Mohan Gummalam and Gabriel Trionfi, who provided assistance in my use of it for studying peer effects.

This dissertation has benefited substantially from the contributions of each of my committee members. Jeremy Bailenson taught me about social influence and persuasion through interactive media. From the very beginning on my work on what would become this dissertation topic, he also fruitfully encouraged me to make this work valuable for wider audiences. Art Owen, through teaching and a collaboration on a statistical article, played an important role in my development as an applied statistician — and in the development of the bootstrap method used in Chapter 4. Byron Reeves provided an excellent introduction to how to think about and work in communication as a field of study, and he more recently helped me make my dissertation work more accessible to communication scholars. David Rogosa introduced me to the fundamentals of applied causal inference. His office was the site of generous sessions discussing encouragement designs and causal inference more generally.

My doctoral studies were supported by grants from Nokia and the National Science Foundation (NSF IIS 0904321). My intellectual development during my doctoral studies benefited from collaboration and discussion with my colleagues at Nokia, especially Tico Ballagas, Jofish Kaye, and Mirjana Spasojevic. This dissertation also benefitted from comments and encouragement by Sinan Aral, Peter Aronow, Dan

McFarland, Paolo Parigi, Cyrus Samii, Cosma Shalizi, Sean Taylor, and audiences at Facebook, the Networks, Histories and Theories of Action seminar and the Research on Algorithms and Incentives in Networks seminar at Stanford, the BIRS workshop on Current Challenges in Statistical Learning, Google Research, and the IMA workshop on User-Centered Modeling.

Finally, to my family and friends, thank you.

Contents

Abstract	iv
Acknowledgements	vi
1 Introduction	1
1.1 Contributions	1
1.2 Overview	3
2 Peer effects	4
2.1 In-kind and out-of-kind peer effects	6
2.2 Epidemiological models: Simple contagion	8
2.3 Economic analysis	10
2.3.1 Graphical games	13
2.3.2 Heterogeneous agents	17
2.4 Psychological analysis	18
2.4.1 Unanimous majority, majority, and minority influence	20
2.4.2 Motives underlying social influence	23
2.5 General discussion	24
3 Research designs for studying peer effects	26
3.1 Causal inference	27
3.1.1 Potential outcomes	27
3.1.2 Structural equations and causal graphs	31
3.1.3 Ignorability and the back-door criterion	31
3.1.4 Front-door criterion	33
3.1.5 Instrumental variables	35
3.2 Adjustment and matching	37
3.3 Experiments	40
3.3.1 Manipulating peer effect mechanisms	41

3.3.2	Confederate peers	44
3.3.3	Manipulating tie formation and network structure	44
3.4	Instrumental variables	47
3.4.1	Friends of friends' characteristics and behaviors as instruments	48
3.4.2	Peer encouragement designs	49
3.5	Conclusion	50
4	Experimental evaluation of observational methods	51
4.1	Original experiment	55
4.2	Method of analysis	58
4.2.1	Multiple behaviors	59
4.2.2	Nonexperimental control group	60
4.2.3	Estimation methods	61
4.2.4	Variable selection and model specification	64
4.2.5	Evaluation	68
4.3	Results	71
4.3.1	Pooled analysis	71
4.3.2	Estimates by domain popularity	75
4.3.3	Estimates for individual domains	80
4.3.4	Explaining differences from experimental estimates	85
4.4	Discussion	87
5	Peer effects in the development of cultural rituals	90
5.1	Method	92
5.1.1	Expressions of gratitude	93
5.2	Model	93
5.2.1	Estimation and dependence	96
5.3	Results	98
5.3.1	Prevalence and effect of prompting individual users	98
5.3.2	Aggregate effect of prompting peers on peers	100
5.3.3	Effects of peer behaviors on egos	104
5.3.4	Observational analysis	106
5.4	Discussion	113
5.4.1	Limitations	114
6	General discussion	116
6.1	Causal inference	117
6.2	Computational social science	118

A	Supplemental information for experimental evaluation of observational methods	120
B	Supplemental information for study of peer effects in cultural rituals	123

List of Tables

4.1	Variables included in models predicting exposure. Note that some of the rows in this table account for multiple inputs to a model. The final column indicates which models include the corresponding variables as predictors.	65
-----	---	----

List of Figures

2.1	Illustration of an exposure–adoption function, in which adoption is caused by aggregate peer behavior. Note that an exposure–adoption function is neither (a) simply a summary of an association or (b) a population-average of a dose-response function.	5
2.2	Illustration of strikingly different exposure–adoption functions that follow from simple and complex contagion models. The simple model here is $1 - (1 - q)(1 - p)^{d_i(t)}$, in which each infected peer can independently cause the ego to adopt and the ego can adopt spontaneously with probability q . The complex model is a case of noisy best response dynamics in a SAGG with strategic complements. Note that these are both individual-level dose–response functions, not population averages.	16
3.1	Graphical causal model with an instrument. D and Y have unspecified joint causes that act through U_D and U_Y . The total effect of Z on Y is exhausted by its indirect effect on Y via D . Thus, Z may be a valid instrument.	36
4.1	Individuals randomly assigned to have exposure to peers sharing a URL prevented are less likely to share that URL. (a) Comparison of probability of sharing in both experimental conditions. Sharing is associated with the number of sharing friends even in the absence of exposure. (b) The average effect of exposure is positive and is larger for individual–URL pairs with more sharing peers. (c) The relative probability of sharing is largest for the case of a single sharing peer. The analysis in this chapter deals exclusively with individual–URL pairs contributing to the first points on the x -axis. [Reproduced from Bakshy et al. (2012b, Figure 4).]	56

4.2	Estimated peer effects via News Feed using stratification on propensity scores. Estimated (a) probability (i.e., risk) of sharing for individual–URLs not exposed (b) risk difference and (c) risk ratio for all qualifying individuals and domains. Experimental (<i>exp</i>) and naive observational estimates are included for reference. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.	72
4.3	Error in estimated peer effects via News Feed using stratification on propensity scores, treating the experimental estimates as the gold standard. (a) Relative error in risk difference estimates for all qualifying URLs. (b) Percent error reduction from the naive observational estimate. Models including same domain shares result in the most error reduction, despite increased variance. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.	74
4.4	Reduction in error in estimated peer effects from adding same domain shares variable to propensity score models. This predictor results in substantial error reduction when treating the experimental estimates as the gold standard. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.	75
4.5	Estimated probability of sharing for unexposed individual–URLs pairs, $P(Y = 1 \mid M = 1, \text{do}(M = 0))$, by the popularity of each URL’s domain. Experimental estimates are superimposed in light blue for reference. Using a propensity score model with the same domain sharing variable substantially increases this estimated probability only for more popular domains, though it increases the sampling variance in all cases. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.	77
4.6	Error in estimated peer effects via News Feed using stratification on propensity scores, treating the experimental estimates as the gold standard. (a) Relative error in risk difference estimates for all qualifying URLs. (b) Percent error reduction from the naive observational estimate. Models including same domain shares result in the most error reduction, despite increased variance. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.	78

4.7	Change in error in estimated peer effects from adding same domain shares variable to the B and S propensity score models by the popularity of each URL’s domain. Values less than zero represent reductions in error. For all categories of domains, adding this variable is estimated to reduce error, though this difference is only significant for more popular domains, and it is notably larger for the most popular domains. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.	79
4.8	Probability of sharing a specific URL in the feed and no feed experimental conditions for the 15 domains with the largest number of exposed individual–URL pairs; numbers in brackets indicate this rank. Domains are sorted by the difference between the two conditions (i.e., the experimental estimate of the risk difference). Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.	81
4.9	Error in estimated peer effects via News Feed using stratification on propensity scores for the 15 domains with the largest number of exposed individual–URL pairs; numbers in brackets indicate this rank. Domains are sorted by largest error. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.	82
4.10	Change in error in estimated peer effects from adding same domain shares variable to the B and S propensity score models for the 15 domains with the largest number of exposed individual–URL pairs; numbers in brackets indicate this rank. Values less than zero represent reductions in error. For most of the domains, adding the predictor results in a smaller estimated error. In some cases, this is statistically significant at the level of individual domains. On the other hand, it also substantially increases it for some domains, generally by underestimating peer effects, as can be seen in Figure A.1. Domains are sorted by the naive error in risk difference. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.	84
5.1	Top left portion of the Facebook home page as of fall 2010. The standard (control) prompt is shown. For users assigned to the alternative prompt ($V_i = 1$), the light grey text instead read, “What are you thankful for?”	92

5.2	Causal DAG representing the relationships among variables in the Thanksgiving Day study. All of the solid black nodes are expected to have common causes U , which are not shown to make the diagram considerably simpler. On the other hand, the instrument Z is colored differently because its only cause, F , is shown. Z is a valid instrument when conditioning on F . The exclusion restriction is encoded by the lack of a path from Z to Y except via D	97
5.3	Probability of ego behavior as a function of observed ego characteristics — gender and previous posting of status updates. Being female and previously posting status updates both predict posting a grateful status update within the first hour after login. Error bars are 95% confidence intervals.	99
5.4	Effect of prompting individual users on occurrence and timing of posting grateful status updates. (a) Cumulative proportion of individuals with the control and alternative prompt posting at least one grateful status update by a number of hours after login. (b) Probability of alternative prompt causing first grateful post to occur prior to some number of hours after login. (c) Ratio of cumulative proportions. Much of the difference in adoption of the behavior between the control and alternative occurs shortly after login, and during the first hour after login (dashed line) in particular. The ECDFs are not exactly zero prior to the login time due to errors in measurement of first login times.	101
5.5	Aggregate peers expressing gratitude, D , by number of prompted peers, Z , and number of peers, F , for the first decile of F . Error bars are 95% confidence intervals and implicitly indicate variation in the number of egos for each value of Z . Some combinations of Z and F with very little data are excluded.	102
5.6	Average linear effect of prompting peers, Z , on peer expressions of gratitude, D . (a) Estimates aggregated by deciles of F , excluding the top decile, and whether the ego has recently posted a status update. The model for each subgroup includes fixed effects of each value of number of peers, F . The effect of prompting an additional peer varies with F . Error bars are 95% confidence intervals. (b) Estimates for each value of F summarized by a local regression weighted by the number of observations.	103

5.7	Instrumental variable estimates of peer effects by deciles of number of peers, F , and prior posting of status updates within the first hour after login. Estimates for some of the subgroups are relatively imprecise both because of varied sample sizes and differences in how much variation in D is caused by Z within that subgroup. Error bars are 95% confidence intervals.	107
5.8	Proportion of egos posting a grateful status in the first hour after login as a function of observed number of peers posting a grateful status update before the ego's login, D . Error bars are 95% confidence intervals.	109
5.9	Proportion of egos posting a grateful status in the first hour after login as a function of observed number of peers posting a grateful status update before the ego's login, D . Each panel is a number of peers who log in an hour before the ego, F , with $F \leq 10$. Error bars are 95% confidence intervals.	110
5.10	Estimated peer effects from a purely observational analysis by deciles of number of peers, F , and prior posting of status updates. This is a OLS regression of Y on D and indicators for each value of F . The indicators for F are centered and interacted with D so as to, under ignorability, estimate the ATE rather than a precision-weighted average. Error bars are 95% confidence intervals.	111
A.1	Estimated peer effects via News Feed using stratification on propensity scores for the 15 domains with the largest number of exposed individual–URL pairs; numbers in brackets indicate this rank. Domains are sorted by the experimental estimate. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.	121
A.2	Estimated peer effects in terms of relative risk for the 15 domains with the largest number of exposed individual–URL pairs; numbers in brackets indicate this rank. Note the very large experimental estimate for www.npr.org	122
B.1	Mean number of peers posting status updates before the ego logs in, D , as a function of observed number of peers, F . Perhaps surprisingly, the relationship is notably linear, given that it might be expected that peers of individuals for (a) log in at different times and (b) have different number of total peers might have systematically heterogeneous patterns of peer behavior. Error bars are 95% confidence intervals.	124

B.2	Probability of ego behavior as a function of observed number of peers F . This association is notably nonlinear, suggesting that care should be taken that Z is made ignorable by appropriately conditioning on F . Error bars are 95% confidence intervals.	125
B.3	Estimated peer effects from a purely observational analysis regressing Y on D and indicators for each value of F . The indicators for F are centered and interacted with D so as to, under ignorability, estimate the ATE rather than a precision-weighted average. Error bars are 95% confidence intervals.	126
B.4	Two-stage least squares estimates are weighted by the variance of the fitted values for D from the first stage. This figure displays factors contributing to the weighting function. The normalized frequency and weights panels are normalized within each decile of F to illustrate the relative weights for each estimate in the IV analysis. Comparison of those two panels illustrates that egos with larger values of F are up-weighted in the TSLS estimation.	127
B.5	Instrumental variable estimates of peer effects by deciles of number of peers, F , and prior posting of status updates within the first hour after login. These estimates are for an alternative model not described in the text. This model includes an interaction of Z with a piecewise linear function of F with knots at the even percentiles of F . This increases the number of instruments and thus weak instrument bias. The results are qualitatively similar to the model reported in the text. Error bars are 95% confidence intervals.	128

Chapter 1

Introduction

The central question of quantitative research on communication has long been, *who says what to whom with what effect?* (cf. Lasswell, 1948). While much of this question is simply descriptive, the causal portion of the question — *with what effect* — has arguably most directly animated research across the social and behavioral sciences. More generally, the social and behavioral sciences have long been concerned with how the behavior of individuals’ is affected by the behavior of those others they interact with and observe. While so central to many theories and contemporary research programs, there are still substantial limitations to our knowledge about peer effects. We at least partially attribute this situation to challenges in credibly learning about peer effects from empirical work.

1.1 Contributions

This dissertation makes methodological and theoretical advances in the study of peer effects. This section anticipates the more detailed presentation of work in each

chapter by outlining the primary contributions.

1. *Analysis of existing research designs.* Recent developments in causal inference have made it increasingly easy and fruitful to give a more formal treatment to causal research questions. This dissertation uses the potential outcomes and graphical frameworks for causal inference to describe the assumptions under which research designs can identify peer effects. This includes a taxonomy of experimental and quasi-experimental strategies. This review highlights challenges for some methods currently seeing increased use, such as natural experiments with random peers and experiments that manipulate mechanisms of peer effects.
2. *Empirical evaluation of observational methods.* While a formal analysis of the assumptions required for a method to identify peer effects is valuable, this does not yet answer how these methods perform with real data, especially considering that critical and untestable assumptions are often violated to some degree. We conduct the first constructed observational study of peer effects, which allows comparison of experimental and observational estimates. Our results here highlight the potential of observational methods to provide informative estimates of peer effects. In particular, estimators using a measure of prior closely-related behavior reduce relative bias to approximately 3%, when treating experimental estimates as the gold standard. In contrast, estimators not making use of this measure exhibit overwhelming bias.
3. *Identifying peer effects in social networks with encouragement designs.* The use of encouragement designs for identifying peer effects in social networks has so far received little attention. We make the case that these methods are

often among the best available for testing for and identifying peer effects. We illustrate this method with a study of cultural rituals.

4. *Peer effects in the development of culture.* How individuals' decisions produce cultural phenomena is a central question for social science. Through analysis of a peer encouragement design and observational data, we provide evidence of positive peer effects in evolving cultural rituals in a study of the behavior of over 70 million people.

1.2 Overview

The remainder of this dissertation is structured as follows. Chapter 2 reviews definitions and theories of peer effects from multiple literatures. This review is organized around the exposure–adoption function, which determines individual's behaviors as a function of exposure to peer behaviors. Chapter 3 introduces frameworks for formal causal inference and systematically reviews research designs used to study peer effects. We observe that there is generally no “gold standard” experiment available and that alternatives have varied and substantial shortcomings for identifying peer effects. However, we suggest that peer encouragement designs are currently underexploited. Chapter 4 analyzes a large experiment as a constructed observational study, allowing for evaluation of the accuracy of observational methods for estimating peer effects. Chapter 5 illustrates the use of a large peer encouragement design. We offer concluding remarks and general discussion in Chapter 6.

Chapter 2

Peer effects

Peer effects are the effects of the behaviors (broadly considered) of an individual's peers (broadly considered) on that individual's behavior. In the standard case, we are interested in *in-kind* peer effects, in which a particular type of behavior by peers causes that same behavior by the focal individual.

In order to make to make this idea slightly more formal, we can consider the non-parametric structural equation¹ for an individual i 's behavior

$$Y_i \leftarrow f_Y(D_i, U_i)$$

where D_i is a vector or scalar measure of the prior behavior of peers and U_i is the vector of all other variables that affect Y_i . This function says that D_i and U_i together causally determine Y_i . We are then most often interested in summaries of the function f_Y , such as the effect of changes in D_i averaged over the values of U_i according to their distribution in a population of interest. Or we may want to

¹This formulation is of a piece with reasoning about causality using graphs (Pearl, 1995, §2.2). The structural/graphical and potential outcomes frameworks for causal inference are presented in Sections 3.1.2 and 3.1.1.

compare the effects of changes in D_i for different values of U_i . Considered somewhat more informally, we refer to this function as the *exposure–adoption function*. That is, this is the dose–response function in which the dose is exposure to peer behavior and the response is adoption of the behavior by the focal individual.²

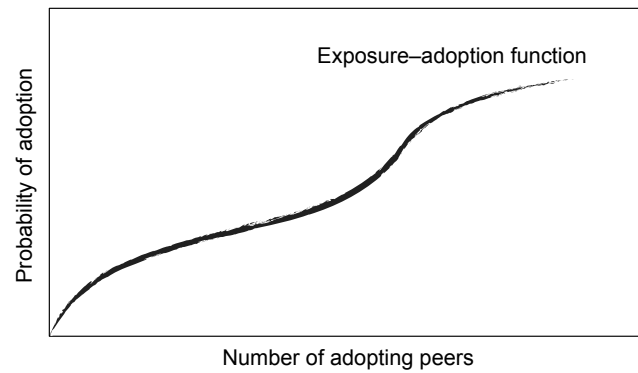


Figure 2.1: Illustration of an exposure–adoption function, in which adoption is caused by aggregate peer behavior. Note that an exposure–adoption function is neither (a) simply a summary of an association or (b) a population-average of a dose-response function.

We take the study of this function as encompassing the central concerns of the study of peer effects. For example, if we wish to understand how culture — the totality of socially transmitted preferences, expectations, and practices — emerges from the local activities of individuals, then we will be interested in individuals’ behaviors as a function of prior peer behaviors. In particular, the clustering of behavior that partially constitutes cultural rituals can result from an upward sloping exposure–adoption function — that is, from positive peer effects.³ Of course, there

²Others refer to this as the influence response function or the contagion function.

³More technically, we instead want to define positive peer effects using the class of weakly

are many research questions that do not fit this mold. In particular, focusing on this function constitutes a focus on questions about the behavior of individuals or, at least, methodological individualism. Thus, those studying questions at the level of the population without providing “microfoundations” for their results may be able to sidestep considering this function in any detail.⁴

2.1 In-kind and out-of-kind peer effects

Much of the literature on peer effects, especially outside of psychology, focuses on what we called in-kind peer effects above: these are cases where the same focal behavior is both the cause and outcome of interest. For example, work has focused on how peer cigarette smoking behavior affects ego smoking (e.g., Christakis and Fowler, 2008; Fletcher, 2010; Lundborg, 2006). Adoption of communication products, which are thus expected to have network effects, has also often been studied in terms of in-kind peer effects (e.g., Aral et al., 2009; Tucker, 2008).

However, in both of these cases, there are reasons to be interested in how the focal behavior is affected by non-focal, perhaps closely related, peer behaviors. Psychologists studying social influence, in particular, have been interested in the effects of how information about the focal peer behaviors are encountered, framed, and summarized (Section 2.4). The innovation diffusion literature has been concerned with the spread of information about an innovation, where often individuals are engaging in this communication behavior of telling others about an innovation without actually adopting the innovation themselves. When a communication product and

monotonically increasing functions that are strictly increasing in some values of D_i .

⁴This includes some research on innovation diffusion, as mentioned in Section 2.3.2. Also see the beginning of Section 3 for associated methodological comments.

service can be used in many ways, researchers and decision-makers may also be interested in which of these uses by peers most affect ego uses (e.g., Burke et al., 2009). Studies of socialization into groups (e.g., Burke et al., 2010) can be understood as the study of both in-kind and out-of-kind peer effects.

This distinction between in-kind and out-of-kind peer effects is similar but not identical to the usual gloss of Manski's (1993) distinction between endogenous effects and contextual effects, which has been influential in the economics literature (e.g., Cohen-Cole and Fletcher, 2008b). *Endogenous effects* are often explained as the result of simultaneous decision-making by individuals about what is, in some sense, the same or parallel decision (Durlauf and Ioannides, 2010, §3); thus the others' focal behaviors (or expected focal behaviors) directly enter into each individuals' decision.⁵ On the other hand, *contextual effects* are the effects of peers' characteristics. These peer characteristics are fruitfully understood as affecting individuals, at least in substantial part, through non-focal peer behaviors; thus, at least much of what is referred to as contextual effects might be fruitfully considered as out-of-kind peer effects. Nonetheless, peer characteristics can affect egos via other processes, such as by affecting the beliefs of non-peers about ego characteristics.⁶

A definition of peer effects should not limit them to the relatively narrow and simplified case of in-kind peer effects. This is especially important as interdisciplinary synthesis in the literature on peer effects brings attention to the study of the mechanisms by which peer effects occur. Aral (2011) also argues for including the out-of-kind case and gives the example of a man encouraging a female friend to

⁵This line of analysis is continued in Section 2.3.

⁶For example, consider how peer race could affect others' expectations about the ego. It seems that the effects of interest are not exhausted by anything we would normally call a peer behavior (cf. Rosenbaum and Rubin, 1984a; Greiner and Rubin, 2010). For a recent study in the context of an online social network, see Walther et al. (2008).

adopt products only for females, such as intrauterine contraceptive devices, because his other female friends use them. Thus, while for simplicity we use the in-kind case as the default in this review, the out-of-kind case is also included conceptually.

2.2 Epidemiological models: Simple contagion

Epidemiologists studying infectious disease have developed models for the infection of individuals as a function of the infection of individuals they have been in contact with. For example, recent work has used the models described in this section to analyze the spread of flu strains. Despite their origin in the study of infectious disease, these models have also been used to model the spread of non-infectious chronic illness (e.g., obesity) and behavior in human and non-human populations. These models are very simple, especially compared with psychological accounts of the affects of many variables on the spread of behavior; they have received attention in part because this simplicity makes them easy to work with. Investigators have also argued that they sometimes provide a useful approximation to real data-generating processes. For these reasons, we begin with these models, despite their lack of grounding in psychological or economic theory.

We can represent the record of the disease-relevant contact between individuals as a time-indexed network g , such that, in discrete time, $g(i, j, t)$ is a measure of the contact between individuals i and j during period t . We will generally take this to be a binary measure and the graph to be undirected, such that $g(i, j, t) \equiv A_{ij}(t) = A_{ji}(t)$ is an entry in the adjacency matrix that specifies the contact network at t .

In the SIRS model, individuals are either *susceptible*, *infected*, or *removed*, and they can progress from each of these stages to next, such that it is possible for

individuals to become infected again.⁷ In discrete time, each infected individual has some probability p of infecting each of its susceptible neighbors in the contact network. That is, the probability of a susceptible individual i becoming infected in period t is

$$P(Y_i(t+1) = 1 | Y_i(t) = 0) = 1 - (1 - p)^{d_i(t)}$$

where we have the number of infected peers $d_i(t) = A_i(t)Y(t)$. We call these *simple contagion* models, since a single infected peer is sufficient for infection and multiple infected peers do not interactively cause infection.

Thus, the probability of becoming infected is a strictly sub-linear function of the number of infected peers. For a small probability of infection p , the function is approximately linear for reasonable numbers of infected neighbors d_i . For larger p , from the perspective of someone hoping to bring about a behavior whose spread is governed by this model, we can say that there are substantial diminishing returns to infecting additional peers. This is an important differences between simple contagion models and models, described below, that are motivated by game theory.

If spontaneous “infection” or “innovation” is possible, then these models can be expanded to include a probability q that the ego becomes infected independently:

$$P(Y_i(t+1) = 1 | Y_i(t) = 0) = 1 - (1 - q)(1 - p)^{d_i(t)}.$$

Studies of diffusion of innovations, rather than spread of infectious disease, often make sure of this larger model.

⁷Other models, such as the SI, SIS, and SIR models can be analyzed as variants of this model. The SIR model was introduced by Kermack and McKendrick (1927). The exposition here follows Jackson (2008, ch. 7) and Centola and Macy (2007).

2.3 Economic analysis

In this section, we review the application of core economic ideas — rational choice theory and noncooperative game theory — to explaining and modeling peer effects. We begin by informally describing peer effects in economic terms; Section 2.3.1 presents the most relevant aspects more formally.

Microeconomic analysis of a phenomena is generally distinctive in its reliance on treating individuals as agents who decide among actions by maximizing the utility they expect to receive (Rubinstein, 2006). This leaves open the empirical questions of whether individuals do in general act in accordance with this model and whether deviations from this model make a substantial difference in outcomes. Since the answers to these questions may vary by domain and “irrationality” is often felicitously modeled by modifications and extensions to this analysis (e.g., Kahneman and Tversky, 1979; cf. Berg and Gigerenzer, 2010), it is worthwhile to work through the predictions of these models.

Thus, an economic analysis of peer effects will characterize them in terms of changes to payoffs and costs, and changes to an agent’s information about — and thus expectations about — these payoffs and costs. We say there are *social interactions* whenever the behavior of an individual is affected by others’ behavior. Economists sometimes make a point of restricting the extension of this term to cases where this happens “directly”, so as to exclude interactions that occur via anonymous markets setting prices (Durlauf and Ioannides, 2010). An influential accounting peer effects in economics and sociology is Manski (2000), which distinguishes three types of social interactions as follows.⁸

⁸Other authors (e.g., Easley and Kleinberg, 2010, ch. 17, 19) simply distinguish between informational effects and direct-benefit effects. These names can be somewhat misleading in light of, e.g., the role knowledge creation can have in constraint interactions, which would be counted as

In *preference interactions*, individuals' preferences depend on the behavior of other individuals. That is, their preference ordering of available actions depends on actions others take — perhaps their underlying costs and payoffs are determined by whether peers take the same action. Individuals' preferences are determined by both genetics and learning; they are endogenous bio-cultural traits (Bowles, 1998). Thus, any economic account of the evolution of culture must consider peer effects resulting from preference interactions.⁹ More narrowly, of interest to us are coordination games in which individuals' prefer to take the same action as others. Preference interactions are the most direct way to incorporate peer effects into games, as considered in Section 2.3.1 below.

In *expectation interactions*, others' behavior changes individuals' expectations about the payoffs for different actions. Individuals may be able to observe the outcome of different courses of action and select their own based on this new information. Or, even without observing outcomes, others' prior decisions can reflect their private information, thus providing novel information to the individual. This distinction is illustrated by the pattern of adoption of hybrid seed corn by farmers (Ryan and Gross, 1943): farmers reported adopting the new seed primarily not after first hearing of others using it, but after hearing about their subsequent crop yields. So at least in this case, we might conclude that observing others' decisions alone provided insufficient information to take this "risky" action.

Note that if all individuals know the true costs and payoffs, then there is nothing to be learned from others' decisions and outcomes, and there will be no expectation interactions, except in predicting others' actions. Likewise, if there is no common direct-benefit effects in this scheme.

⁹Some theorists consider explanation of peer effects via expectation and constrain interactions more desirable than resorting to preference interactions. This is in contrast to the views of Bowles (1998), Manski (2000), and others, who take preference interactions to be pervasive.

variation in costs and payoffs, then others' decisions and outcomes are not informative about costs and payoffs for the ego. In practice, of course, both of these propositions will often not hold, so we can assume that expectation interactions play some role in producing peer effects, if perhaps sometimes a small one.

The final kind of social interaction is most pervasive in economic thinking. *Constraint interactions* are when constraints on resources mean that agents' decisions collectively determine costs and production. Limited production capacity for a good creates constraint interactions among potential purchasers: the more others purchase (and are willing to pay) the less an agent can feasibly purchase; this is a negative constraint interaction. Limited, shared resources also create negative constraint interactions among those who use this resource. For example, decisions to take a particular highway route determine the time costs for that choice. Positive constraint interactions can also occur, such as when public knowledge creation as part of product development increases the the production of both the agent creating the knowledge and others. Some constraint interactions already occur in Arrow–Debreu world of price-setting through anonymous markets, so these are sometimes not considered part of the subject of the literature on social interactions; this limit is sometimes articulated in terms of social interactions generally being externalities (Easley and Kleinberg, 2010, ch. 17; Durlauf and Ioannides, 2010).

This analysis highlights multiple ways that peer effects can emerge, even within the confines of rational choice theory. Note that the shape of the resulting exposure–adoption function can be expected to be quite different for the canonical examples of each type of social interaction. For example, when agents are acting on private information, unboundedly rational choice leads to treating a neighbor's actions as indicators of both the private information that agent has and of the information of

their peers, etc.¹⁰ Furthermore, changes in others' behaviors may reveal that these agents have gained new information, leading individuals to be responsive to rates of change in the prevalence of behaviors (Toelch et al., 2010).

As considered in the next section, preference interactions with only immediate peers can be fruitfully analyzed with graphical games.

2.3.1 Graphical games

Theorists have analyzed peer effects by treating the situation as a game in which the payoffs for each agent and for each possible action depend on the actions of other agents. Standard games of this kind have the payoffs for each agent depend on the actions of all other agents.¹¹ We can generalize this game to cases where each agent's payoffs depend on the actions of some subset of all other agents; these dependences can be formalized as a directed graph such that $A_{ij} = 1$ if and only if agent i 's payoffs depend on the action taken by agent j . For the case where we have an undirected graph and there are only two actions, then $u(a, d, n)$ is the payoff for an agent for taking action $a \in \{0, 1\}$ for whom d of their n neighbors take action 1.¹² This exposition follows Jackson (2008), who calls these *semi-anonymous graphical games* because payoffs only depend on the number of peers taking an action, rather than which specific neighbors take the action; furthermore, all agents with the same degree have the same payoff function.

¹⁰In fact, such a fully Bayesian analysis of expectation interactions is generally intractable and has only been analyzed in limited settings. Analysis of somewhat more nearsighted Bayesian agents is reported by Bala and Goyal (1998) and reviewed in Jackson (2008, §8.2).

¹¹Except as necessary, we do not review noncooperative game theory here. Readers are referred to Osborne and Rubinstein (1994) for an introduction to game theory. Jackson (2008, §9.9) is a succinct introduction to game theory as used in the study of the decisions of individuals embedded in social networks.

¹²Games with more than two actions have been analyzed, but are not treated here.

Given this framework, we can then work out the behavior of agents who are maximizing expected utility and the equilibria of different graphical games. Since we are often interested in situations in which positive peer effects are posited, we focus on games in which the relative payoff for an action increases with the number of neighbors taking that action. In particular, we say that a semi-anonymous graphical game (SAGG) has *strategic complements* if and only if

$$u(a, d, n) - u(a', d, n) \geq u(a, d', n) - u(a', d', n)$$

where a and a' are actions, n is the agent's degree, and $d \geq d'$ are numbers of neighbors taking action a . That is, in these games, the difference in payoffs between any pair of actions increases or stays the same as more neighbors take the former action. Note that the utility is a function of degree; this allows the utility to be linear in the fraction of neighbors taking the same action, for example.

The behavior of an agent who is maximizing expected utility in these SAGGs with strategic complements is exhausted by a threshold: if more than this number of neighbors take action a , then the agent takes action a .¹³ We write this threshold $t(n)$, since it is a function of only the agent's degree.

We mention two relevant special cases. First, say the payoff depends only on the number of adopting neighbors and not on degree — that is,

$$u(a, d, n) = u(a, d, n')$$

for all n and n' . This is a case of *uncontested contagion* because the number of

¹³Agents can be indifferent at the threshold or prefer some particular action. In the case of noisy best-response dynamics considered below, agents might prefer their current action or select actions they are indifferent among with equal probability.

non-adopting neighbors does not affect agents' decisions.¹⁴ On the other hand, if the payoff is determined by, say, the fraction of adopting neighbors, then we have a case of *contested contagion* in which

$$u(a, d, n) \leq u(a, d, n')$$

for all $n \geq n'$.

Consider two ways this SAGG analysis can yield predictions about individual and population-wide outcomes. First, we can identify Nash equilibria for a given game and graph. There are generally multiple equilibria. In the case of the binary action space and strategic complements, there is an equilibrium in which all agents take action 0 and one in which all agents take action 1; there can also be additional equilibria depending on connectivity and other graph characteristics. Thus, from a static analysis, it is often unclear which equilibrium, if any, a system will converge to. Second, we can consider a simple dynamic, turn-based process in which agents make the best response considering only the current actions of their neighbors (Blume, 1995). Other analyses (e.g., Montanari and Saberi, 2010) use noisy best responses dynamics, in which agents make the best response with high probability. These are all considered cases of *complex contagion* (or at least mixed simple and complex contagion) since an adoption requires multiple adopting peers (or at least has a supra-linear increase in probability).

¹⁴This terminology is due to Centola and Macy (2007), who consider both simple and complex contagion models in a common probabilistic framework. Note that simple contagion is a case of (probabilistic) uncontested contagion.

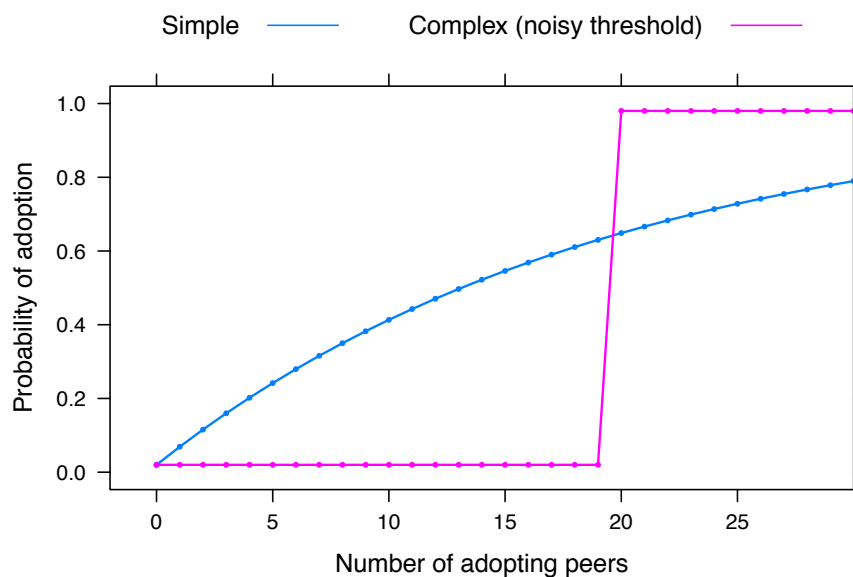


Figure 2.2: Illustration of strikingly different exposure–adoption functions that follow from simple and complex contagion models. The simple model here is $1 - (1 - q)(1 - p)^{d_i(t)}$, in which each infected peer can independently cause the ego to adopt and the ego can adopt spontaneously with probability q . The complex model is a case of noisy best response dynamics in a SAGG with strategic complements. Note that these are both individual-level dose–response functions, not population averages.

2.3.2 Heterogeneous agents

The analysis so far has only allowed for agents to differ from one another in a single way: the utility function $u(\cdot)$ has the agent's degree as an argument. However, we should expect that the payoffs and costs of adoption vary among individuals with the same number of peers. Incorporating heterogeneity into these models makes them more realistic and flexible, but it also makes it more difficult to explore their predictions. Heterogeneous agents are central to current work on peer effects using game theory.

In the games and associated best response dynamics already considered, agents' behaviors are exhausted by a threshold. Therefore, the most direct way to introduce heterogeneity is through agents with the same degree having different thresholds. In the context of a complete network and best response dynamics, Granovetter (1978) developed several results for agents with heterogeneous thresholds. One that is critical for theory and empirical practice is that small changes in the population variance of thresholds can produce qualitative changes in the corresponding equilibrium (cf. Watts, 2002).

While Granovetter's (1978) analysis was motivated by collective action problems such as protests, the model and results are more general. Heterogeneous thresholds are also central to innovation diffusion literature (Rogers, 2003; Valente, 1995; Young, 2009). Since early empirical work with the spread of information about and adoption of agricultural (Ryan and Gross, 1943) and medical (Coleman et al., 1957) products, this literature has used variation in thresholds as one way to account for their observations of particular *S*-shaped adoption curves (Granovetter and Soong, 1983; Valente, 1996).

Heterogeneous agents can be given many other interpretations, such as having variation in culturally determined preferences (e.g., racial prejudice). Schelling (1971) is an early example of analysis based on heterogeneous agents producing striking stylized facts. In particular, there are “tipping points” such that as when racial minority residents reach a threshold in a neighborhood, the racial majority leaves, even if only a small number have strong racial preferences (cf. Card et al., 2008).

We are now in a position to contrast differences in characteristics of exposure–adoption functions that follow from simple and complex contagion (Figure 2.2). Under simple contagion, exposure–adoption functions are sublinear, while under complex contagion, exposure–adoption functions are a step function or at least the largest increases in probability of adoption are for increasing the number of adopting peers to two or more.¹⁵ Thus, application of these two families of models of contagion will generally make quite different predictions about adoption decisions.

2.4 Psychological analysis

Peer effects are a central phenomenon for social psychology. Some of the most striking and influential experiments in psychology have demonstrated the substantial effects the social environment has on individuals’ behaviors (e.g., Asch, 1956; Sherif, 1936). Since the “cognitive revolution” in psychology, these phenomena have been theorized in terms of information processing – invoking multiple information processing systems, numerous heuristics, and associative models of memory (Chaiken and Chen, 1999; Petty and Briol, 2008; Smith and DeCoster, 2000). But compared with cognitive psychology more generally, motives, desires, and emotions have also

¹⁵Note that if the threshold for an agent in a SAGG with strategic complements is a single adopting peer, then the corresponding exposure–adoption function under best response dynamics is also compatible with simple contagion with $p = 1$.

consistently played an important role in psychological theories of influence (Chen et al., 1999; Petty and Briol, 2008).

In this section, we review how psychology has theorized peer effects.¹⁶ On the surface there are some substantial similarities to the economic analysis, but also important differences based in divergent strategies for theorizing human behavior. In particular, social psychology since the 1970s has largely employed a form of *methodological solipsism* (Fodor, 1980), in which variation in external conditions must be specified as variation in stimuli to the individual before this variation can be incorporated into mainstream psychological theory. Thus, in contrast to the study of peer effects in economics, psychologists have been particularly concerned with the specific stimuli that result from others' behaviors and how these are processed. This marks an important distinction between interest in *total peer effects* via potentially numerous mechanisms and interest in peer effects *via a particular mechanism* that is of theoretical interest.

As a result of this concern with the mechanisms of peer effects, psychology has made important contributions to thinking about how investigators should parameterize exposure to peer behavior. In contrast to the best response dynamics in SAGGs described above, this work suggests that models in which only the number or fraction of peers taking an action determines an individuals' response ignore substantial sources of variation in peer effects. Thus, psychological theory is useful for understanding why estimates of peer effects may not generalize across changes in how individuals are exposed to peer behaviors.

On the other hand, this has also meant that what a line of psychological research

¹⁶Communication and consumer researchers studying these topics have primarily worked in this same theoretical space. Other work by communication researchers is discussed elsewhere — in particular, heterogeneous individuals and diffusion of innovations in Section 2.3.2.

says about peer effects is specific to the ways in which individuals' have access to information about peer behaviors. For example, much work on social influence operationalizes norms through statements of descriptive (e.g., that the majority of some comparison group do the focal behavior) and injunctive norms (e.g., that others believe one should do the focal behavior). This may make the results of studies using such stimuli fail to generalize to cases in which peer behaviors are observed or presented in other, perhaps more varied, ways. To the extent that psychology has adopted these stimuli for theoretical reasons or reasons of convenience, rather than because they reflect widespread practices and configurations, this problem may be heightened for application of psychological research.

While graphical games and empirical microeconomics have dealt primarily with behaviors as outcomes (and binary choice in particular), psychological research has often instead focused on attitudes as outcomes. Attitudes are generally measured through behaviors (e.g., circling a number on a questionnaire), but these are often not the behaviors that would be of principal interest to other research communities and to decision-makers. Psychological theory also includes postulates about the link between attitudes and behaviors (e.g., Fazio, 2007; Kraus, 1995). We highlight these differences in focus because they help contextualize differences in the content of economic and psychological theory and in the methods used to develop and test these theories.

2.4.1 Unanimous majority, majority, and minority influence

Within psychological research, the aim of summarizing the shape of the exposure–adoption function is perhaps most clearly addressed by early research on conformity and the continuing literature on persuasion in small groups. This work has studied

how individuals' attitudes and behaviors are affected by the, e.g., decisions and arguments of other members of the same — often small and artificially constructed — group. Reviewing this literature helps to introduce concerns of other work on social influence considered in subsequent sections.

Two early sets of studies are often regarded as launching this line of inquiry.¹⁷ Sherif (1936) used many distinct groups of individuals to examine associations among the judgements made by members of the same constructed group. In a perceptual setting unfamiliar to participants, individuals' perceptual judgements (about the distance a light moved) exhibited substantial statistical dependence with the judgements of others' who expressed their judgements in the same group. Asch (1956) produced substantial changes in perceptual judgements by varying the judgements of copresent others about the same stimuli. In contrast with Sherif (1936) and other earlier studies (e.g., Moore, 1921), Asch used an unambiguous perceptual task in normal lighting conditions — judging which of three lines of substantially different length was the same length as a target line. Participants were faced with a otherwise unanimous majority which, on 12 of the 18 trials, judged that the wrong line was the same length as the target. On these critical trials, 37% of responses conformed with the incorrect majority judgement; 75% of participants conformed for at least one of the 12 critical trials. This study has alternatively been interpreted as a demonstration of independence or of conformity.¹⁸ Whatever the emphasis, the incorrect responses motivated additional theory development: where Sherif's (1936) results could be explained by the ambiguous stimulus conditions and thus reliance on

¹⁷For more on the methodological strategies used in much experimental research with small groups, see Section 3.3.2.

¹⁸For a critical analysis of how American social psychology textbooks have summarized and interpreted these studies, see Friend et al. (1990). In particular, earlier (1953–1964) glosses tended to emphasize independence more than later (1965–1984) ones.

others' judgements for information, this explanation was less convincing in the case of the Asch "conformity" studies. Thus, these results led to Deutsch and Gerard's (1955) development of the informational–normative influence distinction (Section 2.4.2).

An important result of this line of work is that the difference between conformity with a unanimous majority and with a majority one person short of unanimous can be quite substantial. In the Asch paradigm, a single dissenting peer substantially reduced conformity with the majority: under these conditions only 33% of participants made any errors in the direction of the majority. One interpretation of this result is that many individuals have thresholds for compliance in such settings that require that all peers engage in the behavior. Later research, especially in Europe, focused on minority influence in groups (Martin and Hewstone, 2002). For example, Moscovici et al. (1969) demonstrates how a persistent minority can have substantial effects on the attitudes of other group members.

Since then, work on majority and minority influence has been more thoroughly integrated — both by connecting work on majorities and minorities (Martin and Hewstone, 2002), and by reconnecting with the rest of social psychological theory, such as self-categorization theory (Turner et al., 1987) and dual-process models of attitude change (Cacioppo et al., 1986; Petty and Wegener, 1999). In the case of dual-process models, investigators have noted that how influential majorities and minorities are depends on the degree to which their behaviors, judgements, arguments, etc., are elaborated on. For example, minority arguments are elaborated on by other group members when they are consistent with those members' attitudes, but the reverse is true for majority arguments (Baker and Petty, 1994).

2.4.2 Motives underlying social influence

Psychologists have distinguished multiple motives for individuals to change their attitudes and behaviors to accord or discord with others' attitudes and behaviors. The traditional account, following Deutsch and Gerard (1955), distinguishes between informational and normative influence. *Informational influence* occurs when an individual takes evidence of the beliefs and decisions of others as evidence about the truth of those beliefs and the payoffs from those decisions; that is, informational influence relies on an accuracy or efficiency motive. *Normative influence*, on the other hand, involves the motive to act in accordance with social norms; thus, the evidence about others' beliefs and decisions is informative about the norms to be conformed with. Early presentations of this distinction tied this distinction to characteristics of situations and messages that would have their effects via informational or normative influence.

There is a notable surface correspondence between this distinction and two forms of social interactions in the economic analysis reviewed in Section 2.3 above. In both informational influence and expectation interactions, observations of others' behavior are taken as informative about the payoffs from an action, though they do not change the true payoffs. In both normative influence and preference interactions, others' behaviors determine payoffs, making observations of others' behaviors informative about payoffs. However, there are important differences between these taxonomies of processes producing peer effects. Consider cases where there are clear network effects in product adoption. For example, the value of having a fax machine comes from having people to use it to communicate with. Likewise, a complex new product may be more valuable if peers are using it, since they might be able to provide assistance in successfully using it. In both cases, we would analyze this as a preference or, in

the second case, constraint interaction. However, it is not normative influence; in fact, there is no clear place for such cases in this bipartite analysis.¹⁹

Despite the appeal of the informational–normative distinction, which is still widely used, theorists have more recently preferred a three-way taxonomy of motives underlying social influence (e.g., Chaiken et al., 1996; Pool and Schwegler, 2007; Wood, 2000). This family of views distinguishes between (1) motivations to be accurate and efficacious, (2) affiliative motivations to develop and maintain positive relationships with others, and (3) motivations to ensure a coherent and positive self-view. This account expands the motive underlying informational influence, perhaps making any relationship with the economic analysis less clear.

Early tripartite distinctions (Kelman, 1958, 1961) linked these motives to distinct influence processes. The contemporary taxonomies have instead allowed for more flexible relationships between characteristics of messages and situations, motives, and influence processes. Thus, they are generally used in combination with another theory, such a normative focus theory (Kallgren et al., 2000) or a dual-process theory, such as the elaboration-likelihood model (ELM; Cacioppo et al., 1986; Petty and Wegener, 1999).

2.5 General discussion

Since peer effects are of central interest to the social, psychological, and economic sciences, a comprehensive review of the questions, theories, and empirical results that have animated this interest is a challenge beyond the scope of this review. Instead, we have reviewed some of the more influential lines of research with the aim of clarifying

¹⁹Of course, we can offer descriptions of these examples in which normative influence might be involved: if all your peers choose to buy fax machines, you may be violating a norm by refusing to receive faxes. But even without such an elaboration, there is still a preference interaction.

what they have to say about the shape of the exposure–adoption function and its consequences for the spread of behaviors.

Most notably, epidemiology provides models of simple contagion, in which the probability of ego adoption is sublinear in the number of adopting peers, while economic and psychological analyses both suggest ways that many behaviors should exhibit peer effects with more complex exposure–adoption functions. A central question in this area (cf. Centola and Macy, 2007) is which behaviors are fruitfully treated as exhibiting simple contagion (perhaps, e.g., information diffusion) and which instead depend or hugely benefit from adoption by multiple peers (perhaps, e.g., adoption of products with network effects, participation in a protest). Addressing this question requires learning about exposure–adoption functions through difficult empirical work.

Chapter 3

Research designs for studying peer effects

We wish to be able to identify and estimate the probability of an individual engaging in a behavior as a function of the hypothetical behavior of their peers. More modestly, we wish to summarize this exposure–adoption function in some useful way. For many of the models of peer effects under consideration, the behavior of peers enters into this function only as a number of peers adopting; we will treat this as the standard case, though there is reason to be interested in non-anonymous graphical games.

Before turning to identification of the exposure–adoption function, it is worth noting some ways that we can learn about peer effects short of identification of the function or approximations to it. For example, many of the models of peer effects we have considered place restrictions on the shape of observed adoption curves. Here an *adoption curve* is the total number of adopters in a population as a function of time. Young (2009) shows that simple contagion models without spontaneous adoption cannot produce particular *S*-shaped adoption curves; more exactly, they

produce concave adoption curves. Thus, given assumptions about non-peer causes of adoption, the observation of a non-concave S -shaped adoption curve is evidence against the data being produced by those processes.

Setting aside these special cases, the goal of much research on peer effects is the identification of the probability of ego behavior as a function of counterfactual peer behaviors. This is a particularly challenging causal inference problem (Moffitt, 2001). In the following sections, we review the fundamentals of causal inference, conditions under which causal relationships are identified, and how these general results apply to peer effects. On this latter point, we highlight the general implausibility of identification by conditioning on observed variables (cf. Shalizi and Thomas, 2011), as well as the practical and identification problems with experimentation.

3.1 Causal inference

In the next two sections, we review two ways of formalizing causal statements and questions that are suitable for use in statistical analysis. These two formalisms are, as one would hope, equivalent: all theorems in one are theorems in the other (Pearl, 2009b). But they each have advantages for felicitously reasoning about different problems; in fact, both are useful for thinking about peer effects and how to study them.

3.1.1 Potential outcomes

Causal questions can be formulated in terms of what would occur under different hypothetical values of the causes of interest.¹ In our case, the causes of interest are

¹We follow some authors (e.g., Gelman and Hill, 2007, §9.2) in preferring ‘hypothetical’ to ‘counterfactual’ since often one of the values is actual and thus not counterfactual. Other authors use

peer behaviors.

By defining variables for outcomes under these hypotheticals, one can apply the familiar machinery of probability to causal questions. The standard version of this approach and associated notation was developed by Rubin (1974), though some of the core ideas are present in earlier work by Neyman (1923) and Fisher. Thus, this approach is sometimes called the Rubin causal model or the Neyman–Rubin–Holland model. Pearl (1995, 2009b) uses an alternative notation that is more flexible but also more verbose.

Consider the case of a binary treatment D . For each individual, we only observe the outcome Y under a single value of D . Instead of a single variable, we “break apart” Y into two variables $Y^{(1)}$ and $Y^{(0)}$, which are the outcomes that would occur under the treatment ($D = 1$) and control ($D = 0$), respectively; these are the *potential outcomes*. For any individual i , we only observe either $Y_i^{(1)}$ or $Y_i^{(0)}$, but often wish to reach conclusions about their difference

$$\Delta_i = Y_i^{(1)} - Y_i^{(0)}.$$

This casts causal inference as a missing data problem, which Holland (1986) uses to state what he and others (e.g., Gelman and Hill, 2007, §9.2) regard as the fundamental problem of causal inference:

Fundamental Problem of Causal Inference (Holland, 1986). We cannot observe the value of $Y_i^{(1)}$ and $Y_i^{(0)}$ for the same individual i , without changes in other variables, and thus we cannot observe the effect of D_i on Y_i .

In an experiment, individuals are assigned to the treatments at random, so the

‘counterfactual’ (e.g., Morgan and Winship, 2007; Pearl, 2009a). Technical definitions of ‘counterfactual’ avoid this problem, but can still be confusing to readers.

data is missing completely at random. In this case, because of the linearity of expectation, we can identify the average treatment effect (ATE) as:

$$\Delta_{\text{ATE}} = E[Y^{(1)} - Y^{(0)}] = E[Y^{(1)}] - E[Y^{(0)}].$$

This is an important, if modest, result because it demonstrates that we can learn something about the distribution of the Δ_i s even from information about different units.

However, when individuals are assigned to treatments by some other process, it is generally not plausible to assume that the data — the potential outcomes $Y_i^{(1)}$ and $Y_i^{(0)}$ — are missing completely at random: which treatment an individual received, and thus which potential outcome we observe and which is missing, often is caused by the very same factors that cause the outcome. In this case, we say that the treatment assignment is not, in the potential outcomes language, *unconditionally ignorable* or that it is, in econometric language, *endogenous*.² Analysts generally deal with such situations by assuming that the treatment assignment is conditionally ignorable — that the potential outcomes are missing at random given some other observed variables (Section 3.1.3).

Seeing causal inference as a missing data problem leads naturally to the idea of identifiability. Informally, quantity is identified if any change in the quantity corresponds to a change in the probability distribution from which we observe samples. Applied to causal inference with the potential outcomes framework, we can see that identification of quantities like the ATE requires *a priori* restrictions on the probability distribution generating the pattern of missingness in the potential

²These of these terms have been given varying definitions in the literature. See Pearl (2009b, ch. 5) for a critical review.

outcomes; these are causal assumptions. Identification of these causal quantities requires assumptions that restrict the model space.³ Identifiability can be formalized as follows.

Definition of identifiability (Pearl, 2009b). Let $Q(M)$ be a computable quantity of a model M . We say that $Q(M)$ is *identifiable*, given a set of assumptions A , if for any two models M_1 and M_2 that satisfy A , we have

$$P_{M_1}(v) = P_{M_2}(v) \Rightarrow Q(M_1) = Q(M_2)$$

where we observe samples v .

The main aim in the study of peer effects is the identification — and then associated statistical estimation and inference — of quantities summarizing exposure–adoption functions.

Before moving on, we need to introduce another notation that is somewhat more flexible than the standard potential outcomes one. When there are multiple causes of interest, it is difficult to use the potential outcomes notation to refer to the distribution of potential outcomes if the level of only one cause is fixed. For this reason, (Pearl, 2009b) and others sometimes instead use the $\text{do}(\cdot)$ operator, which we introduce here by example: $P(Y \mid \text{do}(D = d))$ is the hypothetical distribution of Y with D held fixed at d .⁴ In this case, we could also express this easily with the potential outcomes notation,

$$P(Y \mid \text{do}(D = d)) = P(Y^{(d)}),$$

³Partial identification — that is, bounding the quantity of interest — can sometimes be an alternative to making additional assumptions (Manski, 2008).

⁴For more on the interpretation of “held fixed” and related ideas of intervention as graphical surgery, see Pearl (2009b).

but in other cases it is more difficult.

3.1.2 Structural equations and causal graphs

While the potential outcomes framework has been widely used by statisticians, econometricians have long made use of the idea of structural equations in reasoning about causality (Haavelmo, 1943). What distinguishes a *structural* equation from other equations is that the sense of equality involved is of Nature *setting* one quantity to be equal to a value determined by others; that is, structural equations are causal claims (Wright, 1921). Thus, the use of ‘=’ is somewhat misleading, and can be helpfully replaced with an operator with a clear direction, such as ‘:=’ or ‘←’ (Pearl, 2009b, ch. 5). For example, if the value of Y is causally determined by variables D , X , and ε , we can write

$$Y \leftarrow f_Y(D, X, \varepsilon).$$

According to this non-parametric structural equation for Y , other variables do not affect Y , except through D , X , or ε . This illustrates that non-parametric structural equations encode qualitative causal assumptions through the exclusion of variables from their right hand sides; these are *exclusion restrictions*.

3.1.3 Ignorability and the back-door criterion

A common identification strategy is to condition on a set of variables X that are arguably sufficient to address any confounding of the causal relationship of interest — the effect of D on Y . That is, investigators make a conditional “no confounding” or “ignorability” assumption. For the case of discrete X , this assumption can be stated using the *do* operator as follows

Conditional ignorability (Pearl, 2009b). Let D , X , and Y be three variables in causal model M . We say that D is ignorable for the effect on Y (or that the effect of D on Y is not confounded) conditional on X if and only if

$$P(Y \mid \text{do}(D = d)) = \sum_{x \in \mathcal{X}} P(Y \mid D = d, X = x)P(X = x).$$

Conditional ignorability is also called conditional no-confounding by Pearl (2009b). Conditional ignorability can also be stated in terms of potential outcomes as

$$\{Y^{(0)}, Y^{(1)}, \dots, Y^{(\bar{D})}\} \perp D \mid X$$

where $D \in \{0, 1, \dots, \bar{D}\}$.

We wish to identify $P(Y \mid \text{do}(D = d))$, and this assumption permits doing so by conditioning on the variables in X and integrating out over their distribution. In practice, instead of this nonparametric conditioning, analysts often use methods such as regression adjustment and propensity score matching that make additional assumptions about functional form but reduce variance of the estimator.

When is conditional ignorability assumption satisfied? That is, when can causal assumptions that follow from substantive knowledge about the data generating process allow us to infer conditional ignorability. We can give a sufficient condition for conditional ignorability in terms of the (perhaps unknown) true causal DAG. The back door criterion gives graphical conditions for conditional ignorability, and thus for identification using adjustment for available covariates X (Pearl, 1995, Theorem 1).

Back-door criterion (Pearl, 1995). Let D , Y , and X be three disjoint subsets of nodes in DAG G . Then X satisfies the *back-door criterion* relative to (D, Y) if for

any pair of nodes (D_i, Y_j) such that $D_i \in D$ and $Y_j \in Y$: (i) no node in X is a descendent of D_i , and (ii) X blocks every path between D_i and Y_j that contains a parent of X_i .

This statement of the back-door criterion depends on defining when a set of variables blocks a path.

Definition of blocking a path (Pearl, 1995). Let D , Y , and X be three disjoint subsets of nodes in DAG G . Let p be a path between a node in D and a node in Y . Then X is said to *block* p if there is a node w on p such that either (i) w has converging arrows along p , and neither w or any of its descendants are in X , or (ii) w does not have converging arrows along p , and w is in X . Here there are converging arrows if any two edges in the path point to the same node in the path.

The back-door criterion formalizes some intuitions that have been widespread among statisticians. First, there is the familiar idea that analysts should condition on the common causes of D and Y . The back-door criterion thus clarifies which common causes must be conditioned on (enough to block all paths). Second, there is the idea that analysts should not condition on “post-treatment” variables — variables that are measured after the treatment begins and that are likely affected by it. This second idea, while common advice from statisticians (e.g., Gelman and Hill, 2007; Rosenbaum, 1984), is still not widely recognized by social scientists.

3.1.4 Front-door criterion

While the back-door criterion provides a general and formal version of informal advice, there is also another sufficient criterion for identification that is less well known. The front-door criterion says that the effect of D on Y is also identified if

we can measure variables that constitute isolated and exhaustive mechanisms for the effect of D on Y .

Front-door criterion (Pearl, 1995). Let D , Y , and M be three disjoint subsets of nodes in DAG G . Then M satisfies the *front-door criterion* relative to (D, Y) if for any pair of nodes (D_i, Y_j) such that $D_i \in D$ and $Y_j \in Y$: (i) all directed paths from D to Y include a node in M , (ii) the back-door criterion is satisfied relative to (D_i, M) , and (iii) D satisfies the back-door criterion relative to (M, Y_j) .

When we have variables that satisfy the front-door criterion, then we can identify the effect of D on Y (Pearl, 1995, Theorem 2). In the case of discrete M , we have

$$P(Y \mid \text{do}(D = d)) = \sum_{m \in \mathcal{M}} P(M = m \mid D = d) \sum_{m' \in \mathcal{M}} P(Y \mid D = d', M = m) P(D = d') \quad (3.1)$$

The interpretation of identification via the front-door criterion is that we can identify the effect of D on Y by seeing how this effect propagates through a mechanism — that is, by identifying the effect of D on M and the effect of M on Y — and then “reconstructing” the effect of D on Y from this partitioning of the joint distribution of D , M , and Y . Note that the front-door criterion thus requires satisfaction of the back-door criterion, with additional restrictions, for both the effect of the treatment on the mechanism, by clause (ii), and the effect of the mechanism on the outcome, by clause (iii).

The front-door criterion is an interesting and novel result of the graphical approach to causal inference. However, it has not yet been widely applied (Morgan and Winship, 2007, ch. 8), likely because many investigators are unaware of it. Additionally, those investigators aware of the front-door criterion may be deterred from

using it in their research by the lack of attention given to the consequences of “minor” violations of it and by lack of training and experience in research designs designed to satisfy it. Some research designs implicitly rely on the front-door criterion (Section 3.3.1).

3.1.5 Instrumental variables

Another strategy for identifying causal parameters is to use instrumental variables. Instrumental variables (or instruments) are, informally, ignorable variables that affect D but do not otherwise affect Y — that is, except through their effect on D (Figure 3.1). The idea is then to see how the variation in D produced by the instrument Z in turn produces variation in Y .

Given the proceeding formalization, readers may already realize instruments cannot offer identification without further assumptions. While the back- and front-door criteria do not fully exhaust the graphical conditions under which the effect of D on Y is identified, instrumental variables strategies generally do not allow identification without making additional assumptions not encoded in the causal graph. Except for some special cases (Balke and Pearl, 1997), identification with instruments generally requires either an assumption of homogeneous effects — that the effect of D on Y is the constant for all units (Holland, 1988) or a shift in focus to average effects of D on Y for an unobservable subpopulation specific to each instrument (Angrist et al., 1996). Given that both the economic and psychological literatures give us many reasons to expect heterogeneous peer effects, the first option is not particularly appealing, except as a working null hypothesis.

The general idea of the second approach is best illustrated with the example of

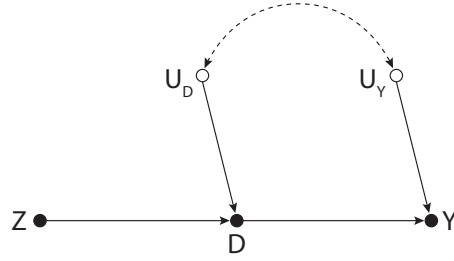


Figure 3.1: Graphical causal model with an instrument. D and Y have unspecified joint causes that act through U_D and U_Y . The total effect of Z on Y is exhausted by its indirect effect on Y via D . Thus, Z may be a valid instrument.

an experiment with non-compliance, which is an example of an *encouragement design* (Holland, 1986). In our example, there are two assignment (or encouragement) conditions and two treatment conditions; that is, both Z and D are binary variables. Thus, $Z_i = 1$ is an encouragement to $D_i = 1$ for individual i . By considering the potential outcomes for the same person i , we can see that there are four types of people. There are the *compliers*, who are caused by the encouragement to take the treatment and do not take it otherwise. There are the *defiers*, who are caused by the encouragement to *not* take the treatment, but do take it otherwise. And there are those individuals whose treatment status is unaffected by the encouragement: *always-takers* and *never-takers*. Note that these types are specific to the encouragement, since one encouragement might induce compliance for a person, while another encouragement would not.

Angrist et al. (1996) develop assumptions under which the encouragement Z allows identification of an average effect of D on Y for compliers — or the *average treatment effect for compliers*. The critical assumption is that there are no defiers.

In the case of a multivalued treatment, this generalizes to the assumption that the causal relationship between Z and D is monotonic. In this more general case, we say that the instrument allows identification of a *local average treatment effect* (LATE), which is a weighted average treatment effect for those units for which D changes in response to Z (Angrist and Imbens, 1995). Since this LATE is instrument specific, its relevance to theory and decision-making is not always clear (Morgan and Winship, 2007; Rosenzweig and Wolpin, 2000). In particular, the clearest case for the LATE being relevant and interpretable can be made when the instrument Z is a randomly assigned and straightforward encouragement of units to the treatment; otherwise, it is often unclear both what variation in D is being used in the analysis and whether Z is actually ignorable (cf. Angrist and Imbens, 1999).

3.2 Adjustment and matching

Having reviewed what is known in general about causal inference, we now return to the case of identifying and estimating peer effects. In this context, D is a variable or set of variables summarizing peer behavior and Y measures the behavior of the ego. This and the next section cover research designs in which investigators assume conditional ignorability, such as by trying to satisfy the back-door criterion.

Adjustment and matching strategies both involve taking some set of available variables X as a sufficient to make D conditionally ignorable. Often we may think of doing this by satisfying the back-door criterion, since this is a clear sufficient condition for conditional ignorability. There is variety in exactly how they condition on these variables, such as through multiple regression, multidimensional matching, or matching or weighting with an estimated propensity score (Morgan and Winship, 2007, ch. 3–5). The assumption that a treatment of interest is conditionally ignorable

is often controversial, but the study of peer effects makes it particularly challenging to make this assumption plausible.

The principal problem is that homophily can also produce the clustering of behaviors that results from positive peer effects (Aral et al., 2009; Manski, 2000). Here *homophily* is the prior clustering of characteristics in a social network; this clustering results from selective tie formation and dissolution and from prior peer effects (McPherson et al., 2001). The term ‘homophily’ (love of the same) suggests a preference on the part of individuals. However, at least within economics and sociology (e.g., Currarini et al., 2010; Kossinets and Watts, 2009), investigators include both individuals preferring to associate with similar others (*preference homophily*) and opportunities for associating with others that are biased towards similar others (*structural homophily*). To understand homophily’s relationship to failing to satisfy conditional ignorability, consider some trait that causes individuals to be more likely to engage in the focal behaviors. If this trait is associated with tie formation or dissolution, which we can reasonably expect it to be, then it is a common cause of ego and peer behavior, creating a back-door path connecting D and Y . If this trait is measured, then an analyst can block that path by conditioning. However, for every such trait that is measured, we can reasonably expect there are many others that are not.

Exposure to external causes of the behavior of the behavior of interest can also be correlated in social networks: peers are also likely to be exposed to the same causes. Investigators sometimes treat this as a source of confounding distinct from homophily (e.g., ‘confounding environmental factors’ in Aral et al., 2009, ‘common external causation’ in Shalizi and Thomas, 2011). In many cases, we can also see this as simply a consequence of prior clustering in whatever traits (e.g., neighborhood, media consumption preferences) that cause exposure to these environmental factors — this

prior clustering being the result of homophily and peer effects in other behaviors. This last point also highlights that in-kind peer effects in a focal behavior can be confounded by in-kind peer effects in another behavior that, in turn, causes the focal behavior.

While there have been several recent high-profile studies that claim to identify peer effects by conditioning (e.g., Christakis and Fowler, 2007), these have appropriately met with substantial skepticism and criticism from statisticians and others (Cohen-Cole and Fletcher, 2008a; Shalizi and Thomas, 2011). The most credible uses of the conditional ignorability, which have used propensity score matching methods to condition on numerous covariates (Aral et al., 2009), highlight the strong assumptions required to regard their estimates as consistent for the peer effects of interest.⁵ One approach here is to use measures of the utility of adopting a behavior as covariates; for example, Aral et al. (2009) use prior use of Yahoo! Sports as a covariate in when estimating peer effects in adoption of a mobile phone application that, among other features, provides sports scores. Another approach, explored in Chapter 4, is to use measures of prior behaviors that are closely related to the focal behavior; however, this may be impossible for, e.g., adoption of a new product, or these measures may simply be unavailable.

We can expect that adjustment and matching estimates of peer effects will continue to have credibility problems, though this may be helped by increased use of sensitivity analysis or analyses of constructed observational studies (LaLonde, 1986), in which observational estimates are compared to experimental estimates; both could

⁵Aral et al. (2009) claim, without proof, that their method is a consistent estimate of an upper bound on peer effects. The idea seems to be that if there is some omitted characteristic that is clustered in the network and predicts adoption, then this would lead to comparing exposed egos who are likely to adopt to unexposed egos who are less likely to be exposed and adopt. The general expectation that these methods should only overestimate (positive) peer effects is applied in Chapter 4.

give the field a sense of the likely magnitude of resulting biases. We conduct the first constructed observational study of peer effects in Chapter 4.

3.3 Experiments

The most straightforward way to satisfy the back-door criterion is to randomly assign individuals to values of D , either unconditionally or conditional on a set of measured variables. Then, in the unconditional random assignment case, we have that

$$P(Y \mid \text{do}(D = d)) = P(Y \mid D = d).$$

This follows from unconditional random assignment because the only parent of D is the mechanism of random number generator, which has no common parents with any variables in the causal graph.

Furthermore, the most conceptually straightforward way to randomly assign individuals to values of D is to *directly manipulate the behaviors of their existing peers*. This is generally not possible; we mention it simply to highlight that it is the generally unavailable ideal. Many of the available strategies are slight deviations from this formula. Experimenters can replace ‘directly manipulate’ with ‘indirectly manipulate’ by assigning peers to an inducement to the behavior of interest (Section 3.4.2). Experimenters can replace ‘the behaviors’ with ‘stimuli constituting exposure to behaviors’ by blocking or modifying a mechanism by which individuals are exposed to their peers’ behaviors (Section 3.3.1). Finally, experimenters can decline to work with existing peers and instead manipulate the behavior of confederates (Section 3.3.2) or randomly assign individuals sets of peers (Section 3.3.3). In what follows, we describe these strategies, highlighting their applications in the literature,

their advantages and disadvantages, and their formal justifications.

3.3.1 Manipulating peer effect mechanisms

As described in Section 2.4 above, investigators in psychology are concerned with how stimuli that convey information about the behaviors of others affect individuals' behaviors. This highlights an opportunity for experimentation: rather than the experimenter manipulating peer behaviors, the experimenter manipulates these stimuli.

To illustrate the idea, consider a simple case where an individual can receive one of two stimuli: the stimulus $M = 0$ provides the information that no peers have adopted the product, while $M = 1$ informs the ego that one or more peers have adopted the product. Let us assume that all peer effects occur by egos responding to this stimulus; this follows from there being no other direct descendants of D in the causal graph besides M . If the experimenter can manipulate this stimulus, then they can identify an average treatment effect for the population for which the stimulus can be manipulated. Sometimes an experimenter might be able to randomly assign individuals to all values of M and thus identify the ATE. In other cases, the experimenter can only block the stimulus when it would have otherwise occurred; that is, the experimenter can randomly assign some individuals who would otherwise have $M = 1$ to $M = 0$. Such an experiment would identify the average treatment effect on the treated (ATET; e.g., the peer effects on the subpopulation who actually has peers adopting the product).

We can fruitfully think of this strategy as an application of the front-door criterion (Section 3.1.4), though investigators usually do not think of it as such. First consider the case where M is an exhaustive mechanism of the effects of D on Y (i.e., satisfying

clause (i) of the front-door criterion). Then we simply need to be able to identify the effect of D on M by conditioning on some available covariates. The idea is that if M is an isolated mechanism then it is only caused by D (and perhaps other random factors not causally related to other variables).

The effect of M on Y , on the other hand, is directly identified by random assignment to M . If M is a deterministic function $h(\cdot)$ of D only, then (3.1) simplifies to

$$P(Y \mid \text{do}(D = d)) = P(Y \mid M = h(d)).$$

Among other possibilities, this can occur if, outside of the experimental intervention, we have that $P(M = D) = 1$. On the other hand, if M is not exhaustive but the other assumptions are satisfied, then the experiment only identifies the effect of D on Y via M . Whether this is an advantage or disadvantage depends on the research question (Morgan and Winship, 2007, ch. 8). For investigators interested in the general shape of the exposure–adoption function, it is likely a disadvantage. If one can assume that peer effects via other mechanisms are nonnegative, then the probability limit of the experimental estimator is a lower bound for the total peer effects.

We are not aware of any experiments of this kind that explicitly make use of the front-door criterion. Instead, investigators simply present their results as estimates of effects of M on Y . This strategy is widely used in the social influence literature in psychology. As we saw in reviewing this literature (Section 2.4), the methodological solipsism of psychology generally motivates theory building around stimuli — and thus specific mechanisms. For this literature, then, identification of only the effect of D on Y via M (or even only the effect of M on Y) is the principal aim. For example, Goldstein et al. (2008) experimented with the signs in hotel rooms encouraging towel reuse by including (false) information about the behavior of other patrons. It is

plausible that these signs are a nearly exhaustive mechanism by which the towel reuse behavior of previous patrons can affect individuals' towel reuse. Even if this is not the case, the primary aim of these experiments was to compare different descriptive norms. Other settings allow for more sophisticated manipulations. In an immersive virtual environment, experimenters can produce rich manipulations of peer behaviors, such as determining the attentiveness of peers in a virtual classroom beset with external distractions (Bailenson et al., 2008).

Some recent work has used this strategy with the aim of identifying peer effects in information sharing behaviors, without reference to mechanisms that have particular importance for psychological theory. Bakshy et al. (2012a) randomly assign some pairs of individuals and advertisements to reduced exposure or non-exposure to social cues in the advertisement that indicate that a peer has expressed affiliation with the advertised brand; they examine resulting differences in responses to the advertisements. This identifies (at least) the effect of these social cues for those pairs that would have have been exposed otherwise; this is a ATET for M on D . Similarly, Bakshy et al. (2012b) randomly assign some pairs of individuals and URLs to a condition that blocks those individuals from seeing their peers share that URL on an online social network service; they examine resulting differences in whether the egos also share that URL. Again, this identifies the effect of this exposure for those pairs who would have been exposed. Neither study makes explicit use of the front-door criterion to attempt to identify the D on Y (e.g., the effect of peer affiliation with the advertised brand), whether in total or via M only.

We say more about Bakshy et al. (2012b) in Chapter 4, where it is the subject of a reanalysis as a constructed observational study.

3.3.2 Confederate peers

Experimenters generally cannot directly manipulate behaviors of existing peers. However, in the case of experiments with small number of individuals who only become the participants' peers for the short duration of, it can be possible.⁶ Often these “peers” are ostensibly fellow participants, but are in fact employed by the investigator as confederates — or, in the language of early work in this area (e.g., Deutsch and Gerard, 1955), “stooges” — to behave according to the instructions of the investigator. Participants are then randomly assigned to observe or interact with confederates who behave according to differing instructions.

From a causal inference perspective, this is perhaps the most straightforward experimental strategy. The disadvantages are also obvious. This strategy can only be used to study the effects of novel peers for short periods of time. It also becomes unworkable when one wishes to investigate the effects of many peers. Finally, individuals' knowledge about their other peers could play a substantial role, e.g., affecting perceptions of the proportion of relevant peers who agree with a policy; this limits the ability of these experiments to address substantial changes in peer attitudes and behaviors already familiar to participants.

3.3.3 Manipulating tie formation and network structure

Investigators can also manipulate who an individual's peers are, apparently addressing confounding due to homophily. Studies that randomly assign individuals to groups can be understood as a simple use of this strategy where each group is a complete graph. Sherif (1936) assigned individuals to groups and measured the degree to which the perceptual judgements of group members were similar; similar strategies

⁶In principle, longer studies are possible, but they are both generally impractical and unethical.

have been used since in the literature on small groups. A larger scale application of this same idea is Salganik et al. (2006): individuals who visited a site for downloading music were randomized to different “worlds” that each separately kept track of the number of times a song had been downloaded. Significant differences among the groups and the distribution of outcomes for particular songs were used to understand the cumulative effects of peer effects on culture. Likewise, Centola (2010) randomly assigns individuals to different network structures when assigning them “health buddies” in an online health intervention.

In most applications of this strategy, the nature of the relationship to the assigned peers is of limited duration, “low-bandwidth”, or even anonymous. Limited duration assigned peers can be compared to confederate peers (Section 3.3.2). Low-bandwidth or anonymous peers can likely only affect the ego via a limited set of mechanisms; such a study has similarities to manipulation of mechanisms (Section 3.3.1). Thus, the most interesting applications of this strategy are when (a) investigators are interested in phenomena beyond individual behavior, such as studying the effects of global network structure on the spread of behavior (e.g., Centola, 2010) or (b) it is possible to assign individuals to peers with whom they will have a lasting and high-bandwidth relationship.

In this latter category, investigators have primarily used variation in the characteristics and behaviors of members of randomly assigned groups and dyads to identify peer effects. Sacerdote (2001) uses random assignment of college roommates to study peer effects in academic performance. Roommate assignment has since been used as an identification strategy by other investigators (e.g., Kremer and Levy, 2008; Zimmerman, 2003). Similar strategies have also been used when groups have exogenous variation in member characteristics and behaviors. Conditional on balancing some member characteristics, new United States Air Force cadets are randomly assigned

to squadrons, which Carrell et al. (2009, 2011a) argue allows them to identify peer effects in both academic achievement and physical fitness.⁷

This strategy has the clear advantage of preventing familiar ways that homophily can be confounded with peer effects (Section 3.2). However, more careful consideration reveals that it requires some strong assumptions in order to estimate peer effects — at least as normally defined. First, if investigators are interested in predicting the effects of interventions in which they change how individuals select their peers, then these experiments have a straightforward interpretation.⁸ But this intervention does not correspond to the normal definition of the counterfactuals of interest: usually we care about what would happen if the behavior of an individual’s existing peers were modified, not what would happen if they were assigned different peers observed to have different behaviors. This distinction lies in the fact that each person is a complex “bundle” of behaviors and characteristics, which often have many back-door paths connecting them.⁹

For example, consider the case of using random squadron assignment to identify peer effects in physical fitness. Let us posit that psychological depression increases the likelihood of failing the physical fitness test. Thus, assignment to a squadron with more cadets who fail the test is also, on average, assignment to a squadron with cadets who are more likely to be depressed. If there are peer effects in depression, then there is a back-door path from peer physical fitness to ego physical fitness. The randomization does not block this path, so it would have to be addressed by

⁷These examples all involve assigning individuals to dyads, groups (i.e., complete subgraphs), or independent networks (i.e., complete or incomplete subgraphs). Treating all of studies as randomization to similarly sized subgraph, we can consider this design generalized to other graphs. An experimenter could randomize people to different clusters in a larger connected graph and examine subsequent clustering of outcomes. We do not know of any studies that implement this design.

⁸Though see Carrell et al. (2011b) for failed application of the authors’ earlier results for designing an intervention of exactly this kind.

⁹This bundle analogy was helpfully suggested by Cosma Shalizi.

assuming ignorability of peer fitness conditional on available variables.

Random peer experiments remain a promising direction for empirical research, but they remain rare opportunities (at least for creating lasting, high-bandwidth ties) and they often do not identify the effects of primary interest — at least without potentially questionable assumptions about additional confounding of multiple peer characteristics and behaviors.

3.4 Instrumental variables

Especially within economics, identification of peer effects with instrumental variables has received growing interest. Proposed instruments can vary substantially in what variation in peer behavior they cause and the plausibility of the associated assumption that the instrument is conditionally ignorable and the exclusion restriction. For example, recent work has used the World Cup to identify peer effects in technology adoption (Tucker, 2008) and surf conditions to identify peer effects in social media use (Shriver et al., 2011). Such idiosyncratic instruments need to be evaluated and interpreted individually. However, there are two more general families of instrumental variables for identifying peer effects, which we review here. The first is a quite general method for identification of peer effects in social networks (Bramouille et al., 2009); it uses the idea that the peers of an ego’s peers only affect the ego through the ego’s peers. The second requires conducting an experiment, but has the advantage of identifying ATEs for an interpretable and relevant subpopulation.

3.4.1 Friends of friends' characteristics and behaviors as instruments

One cause of peer behaviors is the behavior of their peers. Unlike in the study of groups (i.e., complete subgraphs), in most observed social networks an ego's peers have peers who are not mutual peers; that is, they are characterized by some degree of intransitivity. Except for some special cases, we then have that the adjacency matrix \mathbf{A} is linearly independent of the two-step adjacency matrix \mathbf{A}^2 . If we assume, consistent with some peer effect models, that an ego's peers of peers only affect the ego through the ego's peers, then this generates exclusion restrictions for peers of peers behaviors \mathbf{A}^2Y and characteristics \mathbf{A}^2X . Bramoulle et al. (2009) propose identification with these variables as instruments and use it to estimate peer effects in adolescent participation in organized extracurricular activities.

These instruments have the appeal of being available for nearly any observational social network data. Setting aside questions about whether these proposed instruments are in fact conditionally ignorable, there are some challenges of interpretation that may limit the value of the resulting estimates. Unless we are willing to assume, in contradiction of economic and psychological theory, that the peer effects are homogenous, then we must assume monotonicity for IV methods to have a meaningful estimand (Angrist et al., 1996). For illustration, consider the case of using peers of peers' prior behavior as an instrument. Then monotonicity requires that effects of peers of peers' prior behavior has either monotonically increasing or decreasing effect on peers' behavior; that is, we cannot have any egos whose peers have a downward sloping exposure–adoption function, if some others have an upward sloping function. This excludes the possibility of snobbery or anti-conformity (cf. Leibenstein, 1950),

an assumption we may often not want to make in many settings. Furthermore, assuming monotonicity corresponds to assuming positive peer effects, which may be at issue.

Even if investigators are willing to assume monotonicity, the resulting estimates may be difficult to interpret, since they correspond to an average of peer effects based on variation in peers caused by non-mutual peers' behavior or characteristics. It is unclear how this subpopulation is related to a population of interest or to the peer effects.

3.4.2 Peer encouragement designs

One of the more promising developments in identification of peer effects is the use of peer encouragement designs. In a *peer encouragement design*, the investigator assigns individuals to be encouraged to the focal behavior, and examines how this “spills over” to individuals not encouraged. These are sometimes called *partial population experiments* (Moffitt, 2001), and they are seeing increasing use in labor and development economics (e.g., Angelucci and De Giorgi, 2009; Cai, 2011; Duflo and Saez, 2003; Miguel and Kremer, 2004). Here these designs have been primarily used to identify the total effects of assigning peers to experimental conditions, rather than identifying effects of peer behaviors themselves.

If one additionally assumes that the encouragement of an ego's peers only affects the ego through the peers' focal behavior (that is, the exclusion restriction), then summaries of the number of encouraged peers are valid instruments.¹⁰ Duflo and Saez (2003) use this strategy to identify peer effects in enrollment in retirement

¹⁰Since higher degree individuals will have more peers assigned to the encouragement, analysis must condition on degree to satisfy the back-door criterion. An alternative approach may be to use design-based weights in a Horvitz–Thompson estimator (Aronow and Samii, 2011).

accounts. As far as we know, while these designs have been used to identify effects in groups, peer encouragement designs have been neglected for identification of peer effects in social networks. This approach is described further and illustrated with an application to peer effects in the development of cultural rituals in Chapter 5.

We regard these designs as particularly promising for the study of peer effects in online communication behaviors. Companies running large Internet services routinely conduct a large number of experiments testing out new products, minor variations on user interfaces, and modifications to algorithms and statistical models. For example, the size or color of buttons to perform a particular behavior may be modified. Many of these experiments are expected to affect specific behaviors of users. Some of these experiments can then be used as peer encouragement designs for identifying peer effects in those specific behaviors.

3.5 Conclusion

As if often the case in research design, investigators planning empirical studies of peer effects face difficult trade-offs, such as between internal and external validity. Perhaps what sets the study of peer effects — and exposure–adoption functions in particular — apart as more difficult is that credible identification of causal parameters and average effects has remained highly implausible in all but the most artificial or constrained designs. The clever exceptions exploit rare natural experiments, only identify approximations to the average exposure–adoption function, or address substantially more modest questions. Thus, there remains no practical “gold standard” for the study of peer effects.

Chapter 4

Experimental evaluation of observational methods

Strategies for identifying peer effects from observational data require substantial assumptions, many of which cannot be tested with the same data (Shalizi and Thomas, 2011). Homophily is expected to produce positive associations between peer behaviors and ego behaviors that both are not due to peer effects and remain despite conditioning on all available covariates; this would generally result in positive confounding bias such that analysts will overestimate peer effects. Nonetheless, even when these assumptions are not perfectly satisfied, analysis of observational data alone may still produce informative estimates of peer effects. For example, confounding bias may be small compared to the variance of the estimates, such that confounding bias is not a substantial cause of error about the magnitude of peer effects. Empirical and numerical studies can be used to evaluate the performance of these methods with real, or at least realistic, data sets. We call these *method evaluation studies* (Hill, 2008).

There is often substantial uncertainty about what variables should be collected and conditioned on through adjustment, stratification, or matching. Formally, a critical question is whether peer behavior or exposure to peer behavior is conditionally ignorable given some subset of the available variables. In practice, it is often implausible that conditional ignorability is satisfied, but analysts may still aim to substantially reduce confounding bias. How much bias reduction is possible in a particular case is unknown and generally cannot be determined from statistical criteria alone (Pearl, 2009b). Nonetheless, by subjecting observational estimators to method evaluation studies, we can develop informed beliefs about how much remaining confounding bias is expected in observational studies of peer effects. If an observational method results in substantially biased estimates when we can measure the bias, then this speaks against employing such methods when we cannot measure the bias. On the other hand, if a method reduces bias substantially in one setting, then we may reasonably take that method as informative about peer effects under other circumstances in which we believe the causal relationships among the variables are similar.

Method evaluation studies for analysis for observational data can take several forms. Analysts can specify models for selection into treatment conditions and analyze data generated from these simulations (e.g., Hill, 2011; Schafer and Kang, 2008). If multiple similar experimental and observational studies have been conducted, then meta-analysis can compare the results of each (e.g., Heinsman and Shadish, 1996). With a single experiment, analysts can compare experimental estimates of treatment effects to the difference between the outcomes of units assigned to the treatment and others units not in the experiment but simply observed to not have been treated; this is often called a *constructed observational study* (e.g., Heckman et al., 1997; LaLonde, 1986). Similarly, double-randomized studies involve randomly assigning units either

(a) to be randomized to treatment conditions or (b) to the uncontrolled process of selection (e.g., self-selection) into treatment conditions (Shadish et al., 2008).

Except for simulations, all of these method evaluation studies require that at least one experiment has been conducted that directly randomizes the treatment of interest. In the case of total peer effects, this has generally not been possible; that is, experimenters have not been able to assign individuals to different peer behaviors.¹ Thus, while investigators have recognized the value of evaluating observational methods for estimating peer effects, previous work in this area has been limited by this lack of an experimental gold standard. Aral et al. (2009) compare estimates from random matching, as used by Christakis and Fowler (2007, 2008), and propensity score matching, and find that the former produces much larger estimates of peer effects in adoption of a mobile phone application. They offer informal arguments that propensity score matching can only overestimate peer effects, so that this difference demonstrates that the random matching estimator is substantially biased.² However, they are unable to evaluate their propensity-score-based estimators. Cohen-Cole and Fletcher (2008a) apply methods used by Christakis and Fowler (2007, 2008) to estimating peer effects where we should expect none (e.g., height, acne, headaches), and find that these methods nonetheless result in positive estimates of peer effects, which are in some cases statistically significant. Others have used simulations to evaluate performance of observational estimators in the presence of homophily in unmeasured

¹The exception is experiments using confederate peers in small studies (Section 3.3.2), but their small size and lack of homophilous tie formation make them difficult to compare to the large observational studies where the methods of analysis being evaluated are usually applied. This difference in study size is important because it may be that conditional ignorability is only plausibly satisfied, or bias is only substantially reduced, when conditioning on many variables. For example, Aral et al. (2009) include 33 features in each of their logistic regressions predicting exposure. With a small study, this strategy would certainly have large error because of variance in the estimates alone.

²Cohen-Cole and Fletcher (2008b) conduct a similar comparison of methods used by Christakis and Fowler (2007, 2008) with more appropriate techniques.

causes of the behavior (Shalizi and Thomas, 2011; VanderWeele, 2011), when the degree of nodes is censored or ties of varying strength are treated as binary (Thomas and Blitzstein, 2011a,b).

Thus, our understanding of observational methods for studying peer effects — and then of peer effects themselves, can benefit from finding experimental results by which observational results with real data can be judged. It is sometimes possible to randomly assign individuals to exposure to peer behaviors — that is, to manipulate mechanisms of peer effects (Section 3.3.1). Such experiments identify average effects of exposure to peer behavior, which may be of substantive interest in their own right. If the manipulated mechanism is exhaustive (i.e., it is the only mechanism by which the peers behaviors of interest affect the ego behaviors of interest) or nearly so, then such experiments may also enable credibly estimating average total peer effects.

These experiments can be used to create constructed observational studies. Mechanism experiments will generally involve some units being assigned to the same exposure to peer behaviors those units would have otherwise received; that is, the mechanism variable M sometimes assumes its “natural value”, which is also recorded as part of the experiment. An analyst can thus create a constructed observational study simply by excluding from the observational analysis any of the observations where egos are assigned to a different exposure than they would have otherwise received.³ The analyst then uses the remaining observations, where M assumed its natural value, to produce estimates of peer effects using the observational methods to be evaluated. These estimates can then be compared to the experimental ones.

This chapter reports on an analysis of such an experiment. We believe this to

³Such a design has similarities with a double-randomized study. However, if M is only ever set by the experimenter to a single value, as in the study reported on in this chapter, then there is only one level of randomization.

be the first use of a constructed observational study to evaluate observational estimates of peer effects and the largest constructed observational study ever conducted. Specifically, we use stratification on propensity scores to estimate peer effects in sharing links to Web pages (URLs) on Facebook. To anticipate one of the main results, we find that, compared to a naive observational analysis, some of these estimators reduce bias by approximately 80%. This result applies to analysis of peer effects in sharing URLs from relatively popular domain names and when using a measure of prior sharing to this same domain in the propensity score model. The observational analysis also helps us detect some features of the original experiment that cast some doubt on the experimental estimates themselves.

4.1 Original experiment

We reanalyze a large experiment that manipulated an important mechanism of peer effects in information and media sharing behaviors. Bakshy et al. (2012b) randomly assigned some individual–URL pairs to be prevented from being exposed to their peers sharing that URL on Facebook. In particular, for individual–URL pairs assigned to the *no feed* condition, those individuals will not see that URL in their Facebook News Feed or the feed on the profile of peers, whether or not any of their peers have shared it. On the other hand, for individual–URL pairs assigned to the *feed* (i.e., control) condition, those individuals can see that URL and associated comments by their peers; of course, if their peers do not share the URL, they still will not see it. Less than 1% of the over 1.1 billion individual–URL pairs that would have resulted in exposure are assigned to the no feed condition. Even for pairs in the no feed condition, individuals could still see that their peer shared the URL if, e.g., the peer sent it to them in a message or posted it to the individual’s profile. We

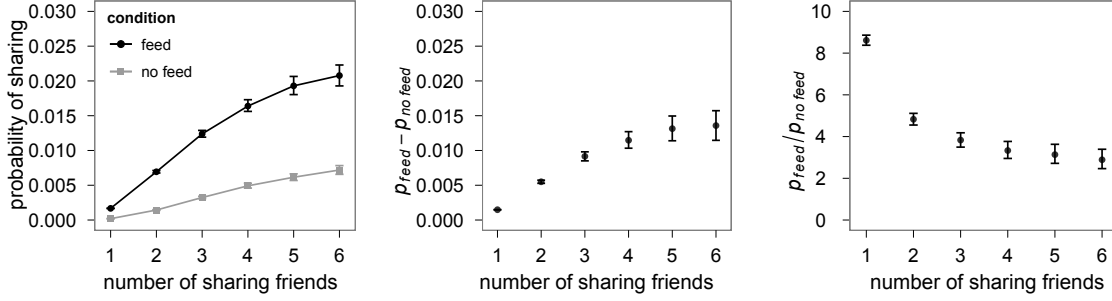


Figure 4.1: Individuals randomly assigned to have exposure to peers sharing a URL prevented are less likely to share that URL. (a) Comparison of probability of sharing in both experimental conditions. Sharing is associated with the number of sharing friends even in the absence of exposure. (b) The average effect of exposure is positive and is larger for individual–URL pairs with more sharing peers. (c) The relative probability of sharing is largest for the case of a single sharing peer. The analysis in this chapter deals exclusively with individual–URL pairs contributing to the first points on the x -axis. [Reproduced from Bakshy et al. (2012b, Figure 4).]

refer readers to Bakshy et al. (2012b) for further details about the experiment and additional analyses of the experimental data.

This experiment identifies the average effect of exposure to peer URL sharing on Facebook for individual–URL pairs for which that individual would have been exposed; this quantity can be described as the average treatment effect on the treated (ATET), for each treatment of exposure to one through six peers sharing the URL. More formally, for each number of sharing peers an individual is exposed to $m \in \mathcal{M}$, the experiment identifies

$$P(Y = 1 \mid M = m) - P(Y = 1 \mid M = m, \text{do}(M = 0)) \quad (4.1)$$

where $Y = 1$ if and only if an individual shares the link in question. If one assumes

that exposure via Facebook News Feed and profile feeds is the exhaustive, deterministic mechanism by which peers sharing a URL on Facebook affects whether the ego shares that behavior, then this experiment would also identify

$$P(Y = 1 \mid D = d) - P(Y = 1 \mid D = d, \text{do}(D = 0)) \quad (4.2)$$

because we always have $M = D$ and thus (4.1) and (4.2) are equal. We can be sure that this assumption is not strictly true. Some individuals can fail to be exposed even when peers share a URL, so the relationship between D and M is stochastic, rather than deterministic. There can also be other ways that peer sharing can affect ego sharing besides exposure; however, it may be the case that, especially for weak ties, exposure via News Feed and profile feeds is almost an exhaustive mechanism of peer effects in URL sharing.

The sample estimates of (4.1) are displayed in Figure 4.1(b). For $m \in \{1, 2, \dots, 6\}$ this provides evidence for positive peer effects in sharing URLs via Facebook. Of particular interest for the subsequent analysis, the estimates of peer effects for the case of a single sharing peer are positive and substantial: exposed individual–URL pairs are 7.37 times more likely to result in sharing that URL than those in the no feed condition.⁴

⁴Note that the estimated value of (4.1) increases with m . It is tempting to interpret this curve as an estimate of the exposure–adoption function or its average over some population, but this would be incorrect. Instead, we interpret this curve as giving estimates of a six different points each from one of six different average exposure–adoption functions, where the average is over six different subpopulations identified by the exposure they would have received without the experiment. While this can be considered a limitation of the original experiment, this is suitable for use as a constructed observational study.

4.2 Method of analysis

We conduct a constructed observational study with this experiment by excluding all individual–URL pairs in the no feed condition from the observational data set. The remaining data is purely observational; that is, these individual–URL pairs have unmanipulated exposure to peers sharing the URL. While the original study includes cases where there are multiple sharing peers $D > 1$ and the ego is exposed to multiple peers sharing the URL $M > 1$, we restrict the present analysis to cases from the original experiment where $M = 1$.

We then augment this data set with a sample of individual–URL pairs observed to have $M = 0$.⁵ This is the full nonexperimental control group. This combined data set is consistent with what data analysts would have if an experiment had not been conducted.⁶ Using this constructed observational data set, we can then produce estimates of ATET (4.1) by using standard methods for estimating effects with observational data. Since we only estimate (4.1) with $m = 1$, we define

$$p^{(0)} = P(Y = 1 \mid M = 1, \text{do}(M = 0))$$

and

$$p^{(1)} = P(Y = 1 \mid M = 1),$$

so that ATET, or the *risk difference*, is

$$\text{ATET} = p^{(1)} - p^{(0)}. \tag{4.3}$$

⁵The population of pairs is very large since it consists of the Cartesian product of all individuals and URLs in the original study, so we use as a sample, as described below.

⁶The only difference is that, since the pairs randomly assigned to the no feed condition are excluded, the set of treated pairs is random sample from the population of treated pairs, rather than the whole population.

This is simply a shorthand for (4.1) with $m = 1$. We then compare observational estimates of (4.3) and other related quantities, such as the risk ratio, $p^{(1)}/p^{(0)}$, to those from the experiment.

4.2.1 Multiple behaviors

While the focal behaviors in this study are all of the same general type (i.e., sharing any URL on Facebook), it is important to recognize that this includes multiple focal behaviors with important differences among them (i.e., sharing one URL versus another URL). Thus, the present study can reasonably be understood as an analysis of peer effects in millions of different, but related, information sharing behaviors.

We expect that there is substantial heterogeneity in what causes individuals to share a *particular* URL. For example, individuals' characteristics that increase the probability of sharing a URL for a news story at `www.nytimes.com` may also decrease the probability of sharing a URL for a discussion of online gaming at `forums.zynga.com`. The experimental analysis can ignore this heterogeneity and still produce unbiased estimates of average effects: the experimental estimates are averages over all URLs weighted by the number of users who would be exposed to their peers sharing these URLs. However, it is expected that an informative observational analyses must address this heterogeneity, since the variables that cause sharing particular URLs are correlated in the network because of, e.g., homophily. The following discussion applies to both regression adjustment in the model for the outcome and propensity score methods, in which the analyst fits a model predicting exposure to peer sharing and then uses the fitted propensity scores in subsequent estimation of peer effects.

One way to address this multiplicity of focal behaviors is to allow the outcome

model or propensity score model to vary by URL or by groups of URLs, such as those that share the same domain name (e.g., `news.yahoo.com`). As described further below, we employ the latter approach. We can only conduct this analysis for URLs or domains with a sufficient number of observations to allow us to fit a model specific to those observations.⁷

4.2.2 Nonexperimental control group

We constructed a nonexperimental control group⁸ with approximately the number of individual–URL pairs in the experimental data set.⁹ The full nonexperimental control group is constructed so as to have a similar marginal distribution of individuals and URLs as the exposed group. That is, URLs appear in the nonexperimental control group the same number of times each appears in the experimental data set. To form individual–URL pairs from this set of repeated URLs, individuals were then sampled with probability proportional to the number of times they appear in the experiment. In expectation, this procedure produces a nonexperimental control group with users and URLs with the same characteristics as the exposed group.

We restrict our analysis to the 3,704 domains with at least 10,000 individual–URL pairs in the exposed and nonexperimental control group combined. We additionally restrict our analysis to Facebook users believed to be located in the United States

⁷Some ways of introducing such a sample size restriction can be regarded as conditioning on the dependent variable. In particular, if one restricts the analysis to URLs for which a large number of users were exposed to a peer sharing that URL, then this is also a restriction of the analysis to URLs that *were shared* by a large number of users. One would then be examining only those behaviors that become more widespread — perhaps because they went “viral”.

⁸In the context of methods in which individual treated and control units are matched with each other, this is sometimes called a *reservoir*.

⁹This is approximate because, for computational reasons, the sampling method used waited until the final step to filter out pairs that were actually exposed. There could be additional gains in precision from using a larger nonexperimental control group, but this would introduce additional computational challenges by increasing the size of an already very large data set.

and using the site in American English. This restriction is designed to eliminate one major source of variation in the exposure and outcome that can be difficult to adjust for. It also makes our more granular analyses by the popularity of the domain of each URL more readily interpretable. This results in an experimental data set with 58 million individuals, 11 million URLs, and 212 million individual–URL pairs. Of these, there are 35 million individuals, 7 million URLs, and 71 million pairs are exposed to a peer sharing the URL; this is the treated group used in both the experimental and observational analyses. The full nonexperimental control group includes 47 million individuals, 14 million URLs, and 142 million individual–URL pairs.¹⁰

4.2.3 Estimation methods

The primary estimand in the original experiment is the average peer effect via News Feed on those who would have been exposed (i.e., ATET). Effects of exposure are likely to vary between those who would have been exposed and those who would not have been. For example, individuals who would have been exposed to a peer sharing a particular news story at `www.cnn.com` might be more likely to be caused to share that URL by the exposure than would individuals who would not have been exposed (e.g., because none of their peers shared it). We regard ATET as of more interest than the average treatment effect on those who would not have been exposed, and this is the quantity for which we have an experimental estimate.

Thus, if an analyst were estimating this average peer effect according to the state-of-the-art in observational methods, they would likely use a weighting, matching, or

¹⁰The restrictions of the analysis here to Americans and popular domains result in the nonexperimental control group being somewhat smaller than the experimental one, since some pairs are filtered out after its construction.

stratification scheme such that resulting estimates of average effects are, under some assumptions, approximately unbiased for the averages for each of the subpopulations.¹¹ We describe the observational estimators used in this and the next section.

Note that both experimental and observational analyses use the same individual–URL pairs to estimate the probability of sharing when exposed, $p^{(1)}$. It is the estimation of the probability of sharing when not exposed, $p^{(0)}$, that varies between the experimental and observational estimators and among the different observational estimators.

Stratification on estimated propensity scores and domain

The primary method used is granular stratification on estimated propensity scores and domain (Rosenbaum and Rubin, 1983, 1984b; Rubin, 1997). The unknown true propensity score

$$e(X_{iu}) = P(M_{iu} = 1 \mid X_{iu})$$

is the probability of an individual i being exposed to URL u , where X_{iu} are measured variables describing that individual–URL pair. We estimate the propensity scores using logistic regression on these variables

$$\hat{e}(\mathbf{X}) = \text{logit}^{-1}(\mathbf{X}\hat{\beta}).$$

We fit a series of different specifications of this model, as described in Section 4.2.4. Since the true model for peer and ego behavior is expected to be heterogeneous across very different URLs, we fit a separate model for each domain, which is equivalent to

¹¹Thus, we can expect that using regression adjustment alone would both be inadequate and not the choice among analysts familiar with contemporary methods, since with heterogeneous effects regression adjustment would result in precision-weighted estimates.

a model in which \mathbf{X} includes indicators for domains and products of these indicators and all other inputs.¹² The resulting estimated propensity scores can then be used in three closely related ways — to construct weights for each unit, to match exposed and unexposed units, or to divide the sample into strata or subclasses. We use stratification (subclassification) on the estimated propensity scores. Such stratification can also be regarded as form of nonparametric weighting or a form of matching, sometimes called “blocking” (Imbens, 2004) or “interval matching” (Morgan and Harding, 2006), that does not impose a particular ratio of treated to control units, as one-to-one matching methods do. For very large data sets, such as the current study, stratification has computational advantages over one-to-one matching, and the larger sample sizes afford using more strata.¹³

To construct the strata, we compute the percentiles of the estimated propensity scores for each individual–URL pair within each domain. These percentiles are then used as the boundaries for stratification of pairs for each domain. The estimates of $p^{(0)}$ for a particular domains is an average of the estimates for each strata weighted by the number of exposed cases within that strata:

$$\hat{p}^{(0)} = \frac{1}{\sum_{i,u} M_{iu}} \sum_{l=1}^{100} \left(q_{iul} Y_{iu} \sum_{i,u} M_{iu} q_{iul} \right)$$

where q_{iul} is an indicator of whether the estimated propensity score for individual i and URL u is in the l th percentile. Estimates from multiple domains are combined in the same way by weighting the estimate for each domain by the number of exposed

¹²For some of the domains, we downsampled the number of pairs somewhat so as to make repeated fitting of these models, as required by our bootstrapping procedure (Section 4.2.5), more practical.

¹³Rosenbaum and Rubin’s (1984b) original presentation of stratification on estimated propensity scores used quintiles, as have many applications since. As the precision of the estimated propensity scores increase (i.e., with more observations), there can be substantial within-strata confounding (Lunceford et al., 2004). This motivates our use of a larger number of strata.

pairs for that domain.

Naive analysis

For the sake of comparison, we also conduct a more basic analysis that does not utilize the propensity scores or other adjustment. To estimate the probability of sharing for unexposed individual–URL pairs, we simply compute the proportion of individual–URL pairs in the nonexperimental control group that shared the URL for each domain. For analyses of multiple domains, we average these estimates, weighting each by the number of exposed individual–URL pairs for that domain. Because the method by which the nonexperimental group was constructed approximated the marginal distribution of users from the exposed group, this approach can be seen as finding unexposed individuals similar to the exposed individuals, but without any adjustment for propensity to be exposed to different URLs. In the subsequent analysis, we refer to the resulting estimates as the *naive observational estimates*.

4.2.4 Variable selection and model specification

There are numerous variables available for the propensity score model.¹⁴ It is not possible or desirable to include all the variables that an analyst could construct because of the work involved in defining variables, the costs of increasing dimensionality for precision, and computational challenges in using all of them in an analysis. Furthermore, other situations may require that investigators decide in advance what variables are worth measuring. In both of these settings, it is standard practice to use theory and other domain knowledge to select variables. In the case of peer effects

¹⁴For example, an analyst could construct the *individual–term matrix* counting all the words used by each individual in their Facebook communications; each of these thousands of variables could be used as a covariate.

Table 4.1: Variables included in models predicting exposure. Note that some of the rows in this table account for multiple inputs to a model. The final column indicates which models include the corresponding variables as predictors.

Category	Name	Description	Models
Demographics	Age	As indicated on profile	A, B, B_s, D
	Gender	As indicated on profile	A, B, B_s, D
	Relationship status	As indicated on profile	A, D
	Political affiliation	As indicated on profile	A, D
Facebook	Friend count	Number of friendships at start of study	A, B, B_s
	Initiation count and prop.	Number and proportion of extant friendships initiated	A
	Tenure	Days since registration of account	A, B, B_s
	Profile picture	Whether the user has a profile picture	A
	Visitation freq.	Days active in prior 30 day period	A, B, B_s
		Days active in prior 91 and 182 day periods	A
Communication	Action count	Number of posts (including URLs), comments, and likes in a one month period	A
	Post count	Number of posts (including URLs) in a one month period	A
	Comment count	Number of comments on posts in a one month period	A, B, B_s
	Like count	Number of posts and comments “liked” in a one month period	A
URL sharing	Shares	Number of URLs shared in a one month period	A, B, B_s, S, S_s
	Unique domains	Number of unique domains of URLs shared in a six month period	A, B, B_s, S, S_s
	Same domain shares	Number of URLs shared in a six month period with the same domain as outcome	A, B_s, S_s

in URL sharing, the analyst would select variables believed to be related to causes of sharing a URL and to be associated with network structure (i.e., peer and ego variables are associated because of homophily, common external causes, and prior influence).¹⁵ Table 4.1 lists the variables we computed based on these expectations. These variables are each included in at least one of the following six models specifications, which are designed to correspond to selections of variables that an analyst might make.

- Model *A* includes all of the variables. This model is expected to have the largest potential for bias reduction but to also suffer from increased sampling variance. In other settings, many of these variables might not be available to analysts.
- Model *B* includes a smaller selection of variables that still spans multiple categories. This selection was determined by consulting with one of the authors of the original experiment and asking him, based on his exploratory analysis of the experimental data only, what individual-level variables he expected to be (a) most predictive of sharing and (b) most likely to substantially reduce bias. This model thus reflects use of substantial domain expertise for dimensionality reduction.
- Model *D* includes demographic variables only. At least some of these variables, or similar measurements, would likely be available in many other settings. These are all expected to be associated with consuming content from particular sources. *D* can also be seen as a convenience selection of covariates.

¹⁵This is not to say that the analysis must think each variable is a likely cause of sharing behaviors, but simply that they are causes of sharing behaviors *or* are decendents of these causes.

- Model S includes only variables that measure the general (i.e., individual-level) behavior of sharing URLs on Facebook. Individuals who previously have shared many URLs and URLs from many different domains are expected to be more likely to share URLs, and, because of homophily and prior influence, to be more likely to be exposed to URLs.

Of the models listed above, only A includes the number of URLs from the same domain shared by the individual in the prior six months. Since this *same domain shares* variable is expected to be one of the best predictors of both exposure and sharing, we also specify two additional models, Bs and Ss that add same domain shares to B and S . This allows for straightforward evaluation of the consequences of using this variable to the observational analysis. We regard same domain shares as an example of more specific information about related prior behaviors. In some cases, such information will be available to analysts. In other cases, this information may not be available, or the related behaviors may not be sufficiently common to be useful.

For all six models, the logistic regression model is specified with indicators for each level of categorical variables (e.g., gender is entered with an indicator each for male, female, and unknown), the logarithm of variables that count actions taken in a fixed period (e.g., comment count), and untransformed versions of all other numeric variables (e.g., tenure). Again, the coefficient for each of these inputs varies by the domain name of the URL (i.e., there are domain indicators that interact with all variables).

Throughout, when displaying the different estimators of peer effects, we group the estimates into four categories: the experimental estimate (shown in blue), the naive estimate (grey), the estimates from propensity stratification that do not make

use of same domain shares (yellow), and the estimates that do make use of same domain shares (pink).

4.2.5 Evaluation

We compute the discrepancy between each of the resulting seven observational estimates and the experimental estimates. Our focus is on estimates of the average treatment effect on the treated (ATET), which is the difference in probabilities of sharing caused by exposure for those individuals who would have been exposed; this is also called the (causal) risk difference. For each estimator $\widehat{\text{ATET}}_k$, we can characterize these discrepancies in multiple ways. Most simply, we can compute the discrepancy on the probability scale; that is, the *error in risk difference*:

$$\widehat{\text{ATET}}_k - \widehat{\text{ATET}}_{exp} = \left(\hat{p}_k^{(1)} - \hat{p}_k^{(0)} \right) - \left(\hat{p}_{exp}^{(1)} - \hat{p}_{exp}^{(0)} \right).$$

To state the size of this discrepancy in a way that account for the size of the experimental estimate, we also report the *relative error in risk difference*

$$\frac{\widehat{\text{ATET}}_k - \widehat{\text{ATET}}_{exp}}{|\widehat{\text{ATET}}_{exp}|}.$$

Finally, since the error of the naive estimate can be very large compared to the experimental estimate itself, we also use a measure of bias reduction: the *percent error change from the naive estimate*,

$$100 \left(1 - \frac{|\widehat{\text{ATET}}_k - \widehat{\text{ATET}}_{exp}|}{|\widehat{\text{ATET}}_{naive} - \widehat{\text{ATET}}_{exp}|} \right).$$

In some additional comparisons, we also use an different estimate as the baseline for this measure of bias reduction.

For much of the analysis, we take the experimental estimates as the gold standard — as unbiased for the causal estimand of interest. This motivates the description of the above quantities as “error”. In additional analyses, we consider whether the experimental estimates should be given this role and whether, in some case, the observational estimates ought to carry equal evidential weight.

Confidence intervals

Our observations of both exposure and sharing are not independent and identically distributed (IID). Individuals vary in their probabilities of exposure and sharing, as do URLs. Exposure and sharing events are dependent, since an individual using Facebook at a particular time can often result in exposure to multiple URLs, and one person sharing a URL affects others’ exposure status. Methods for computing confidence intervals that neglect this dependence structure are expected to be hugely anti-conservative; that is, they would substantially overstate our confidence about the probability limit of each estimator.

To address this issue, all statistical inference in this paper employs a bootstrap strategy for data with this crossed structure (Brennan et al., 1987; Owen, 2007; Owen and Eckles, 2012).¹⁶ We now briefly describe our use of this strategy so that readers can interpret our results.

For each of $R = 50$ bootstrap replicates, we reweight observations according to the following procedure (Owen and Eckles, 2012).¹⁷ For the r th replicate, each

¹⁶This is not true of the original experimental results, reproduced in Figure 4.1, since Bakshy et al. (2012b) only reweight URLs, not users.

¹⁷Software implementing this multiway bootstrap in R is available at https://github.com/deaneckles/multiway_bootstrap. The results in this chapter were produced by similar software

individual is assigned a $\text{Poisson}(1)$ draw, and each URL is assigned a $\text{Poisson}(1)$ draw. Each individual–URL pair is then assigned the product of the corresponding draws as its weight. All procedures are applied to the original data set and each of the replicates, such that each propensity score model is fit $R + 1 = 51$ times, percentiles of estimated propensity scores for each domain are computed 51 times, etc. Under very general conditions, this strategy is known to be conservative when estimating the variance of means (Owen, 2007; Owen and Eckles, 2012). Throughout, we report 95% bootstrap percentile confidence intervals, which are expected to have at least 95% coverage.¹⁸ The bootstrap distribution of matching and resulting estimates is generally biased, since the matches within a bootstrap replicate will be worse than the matches available in the full sample (Abadie and Imbens, 2008). This would increase the bias of the propensity score estimators — thus making our evaluation of these estimators more negative, though we expect that this problem may be reduced by our use of stratification, rather than one-to-one matching.

Note that all of the comparisons of interest are not entirely between-units. For example, the observational and experimental estimates share individuals, URLs, and even individual–URL pairs. Thus, observing that confidence intervals for two quantities overlap does not indicate that their difference (or ratio) is not statistically significantly different from zero (or one). This is one reason why we include figures showing estimates and intervals for these differences and ratios themselves.

for Apache Hive.

¹⁸The above discussion of our bootstrapping strategy is derived from text written for Bakshy et al. (2012a).

4.3 Results

This section reports the observational estimates of peer effects and compares them to the experimental estimates. We begin with estimates for the whole subpopulation. Since both peer effects and the performance of the estimators is expected to vary with the popularity of the domain, we conduct further analyses for subgroups of domains by their popularity, and for individual popular domains. In general and as expected, the observational estimates overestimate peer effects when the experimental estimates are taken as the gold standard. Estimates of peer effects using stratification on propensity scores come very close to the experimental estimates when (a) same domain shares is used in the propensity score model and (b) for subpopulations that include many or only observations of individual–URL pairs from popular domains. Nonetheless, there are some popular domains for which these observational estimates are quite different from the experimental ones, including apparently *underestimating* peer effects. At least in some cases, we find some reasons to doubt that the experiment identifies the relevant peer effects for these domains.

4.3.1 Pooled analysis

Our primary analysis is of estimated peer effects via News Feed for all qualifying domains. We are working with a subpopulation of the individual–URL pairs used in the original experiment, so we first consider the experimental estimates for this population. Comparison of individual–URL pairs in the feed and no feed conditions indicates that the exposure via News Feed causes an increase of 0.00112 (95% CI = [0.00109, 0.00114]) in the probability of sharing that URL. This is similar to, but smaller than, the estimate of an increase of 0.00166 from the analysis of all the individual–URL pairs in Bakshy et al. (2012b). Likewise, we find that pairs in

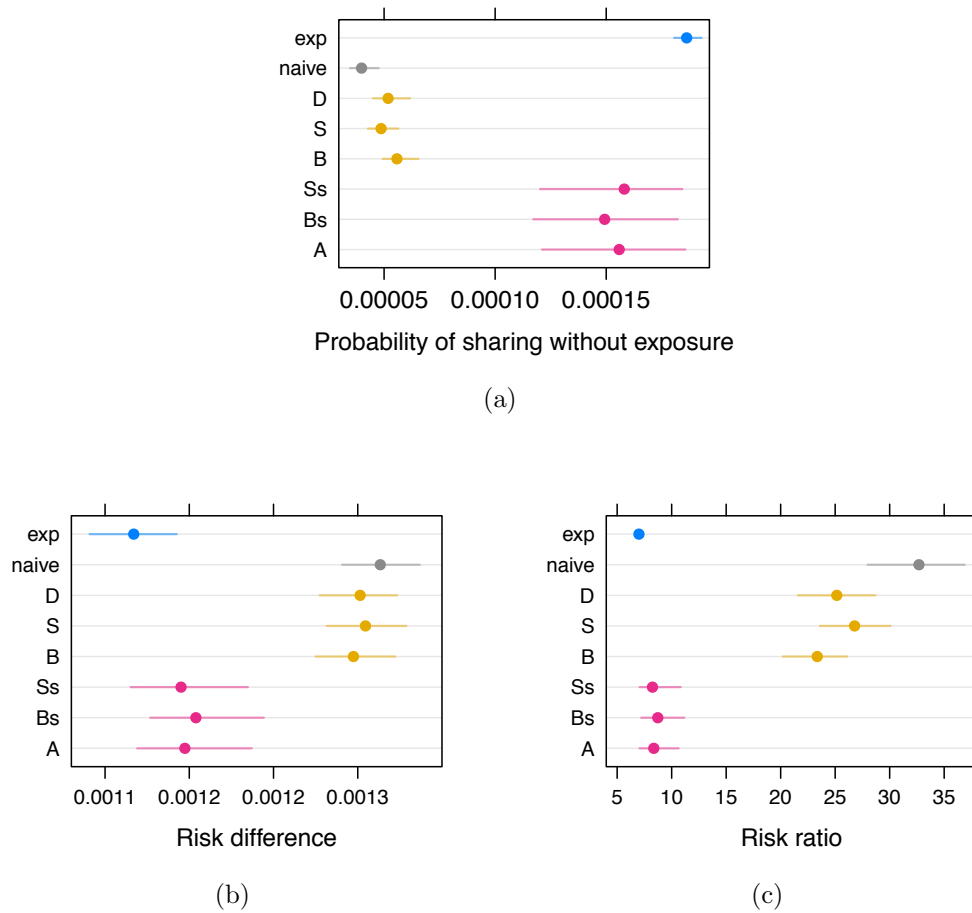


Figure 4.2: Estimated peer effects via News Feed using stratification on propensity scores. Estimated (a) probability (i.e., risk) of sharing for individual-URLs not exposed (b) risk difference and (c) risk ratio for all qualifying individuals and domains. Experimental (*exp*) and naive observational estimates are included for reference. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.

the feed condition are 6.00 (95% CI = [5.77, 6.19]) times more likely to share than pairs in the no feed condition. This is smaller than the full-population estimate of 7.37. These are the experimental estimates of peer effects to which we compare our observational estimates.

Both the experimental and observational estimates use the same individual–URL pairs for estimating the probability of sharing when exposed, $p^{(1)}$, and do not weight these pairs differently. Thus, all differences in estimated peer effects are the result of differences in estimates of the probability of sharing when unexposed, $p^{(0)}$. Figure 4.2(a) shows these probabilities according to each of the six propensity score models, the naive estimator, and the experimental estimator. As expected, the naive analysis of the full nonexperimental control group produces a comparatively small estimated probability of sharing, since many of these pairs are unlikely to be exposed or to share.

These differences among the observational estimates of $p^{(0)}$ carry over to the estimated peer effects. Figure 4.2(b) shows the estimates of the effect of exposure to a peer sharing a URL on the individual sharing that URL. While we focus on the risk difference (since this is the ATET), the differences among the estimates are most striking when cast in terms of relative increase in the probability of sharing. Figure 4.2(b) displays these estimated risk ratios. At the extreme, the naive analysis suggests that individuals are 31.69 (95% CI = [26.97, 35.89]) times more likely to share if exposed than if prevented from being exposed. The estimates from propensity score models that include the same domain sharing variable are notably close to the experimental estimates of the risk difference and risk ratio.

If we treat the experimental estimates as the gold standard, we can characterize the discrepancy of the observational estimates in terms of error (Figure 4.3). The naive estimator $\widehat{\text{ATET}}_{naive}$ has a relative error from the experimental estimate of

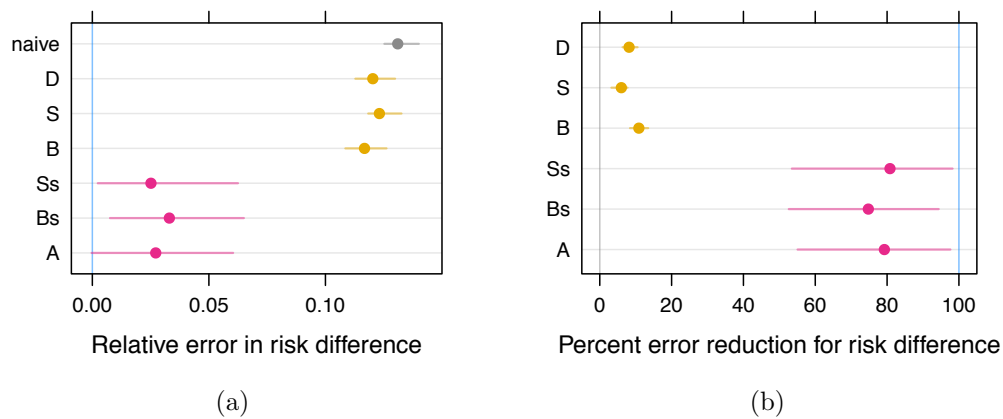


Figure 4.3: Error in estimated peer effects via News Feed using stratification on propensity scores, treating the experimental estimates as the gold standard. (a) Relative error in risk difference estimates for all qualifying URLs. (b) Percent error reduction from the naive observational estimate. Models including same domain shares result in the most error reduction, despite increased variance. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.

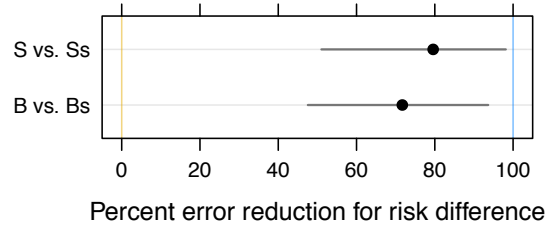


Figure 4.4: Reduction in error in estimated peer effects from adding same domain shares variable to propensity score models. This predictor results in substantial error reduction when treating the experimental estimates as the gold standard. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.

0.131(95% CI = [0.125, 0.140]); that is, its estimate is 13% larger than the experimental estimate. On the other hand, estimates that make use of same domain sharing have much smaller relative errors: their difference from the experimental estimates is approximately 80% smaller than the naive estimates. This is a substantial amount of bias reduction: the remaining bias in these estimates of the risk difference is under 4% of the experimental estimate — and only about twice as large as their standard errors.

4.3.2 Estimates by domain popularity

The preceding analysis provides evidence for the importance of including measures of prior behaviors similar to the focal behavior. In particular, we found that including the same domain sharing variable substantially reduced error in estimating peer effects, taking the experimental estimates as the gold standard. However, it is expected that the effectiveness of this strategy depends on being able to measure prior behaviors that are (a) closely associated with causes of exposure and ego behavior and (b)

relatively common, so as to afford precise estimation. So to better understand how such measures can reduce error of observational estimates, we repeat much of the analysis of the previous section on subgroups defined by the popularity of the URL’s domain, where popularity is measured by the number of exposed individual–URL pairs in the experiment. We expect that the models including same domain sharing will show the most improvement for more popular domains.

Again, since differences among the estimates of peer effects are determined only by differences in the estimates of the probability of sharing without exposure, $p^{(0)}$, we analyze these estimates for each of six subgroups of domains. These subgroups are defined by the 10, 25, 50, 75, and 90% quantiles of the distribution of the number of exposed individual–URL pairs. Figure 4.5 presents these probabilities for each of the estimators, with the experimental estimates overlaid for reference. Propensity score models including same domain sharing produce estimates with notably larger sampling variance for all subgroups, as can be seen from the width of the confidence intervals. For domains in the top quartile (i.e., the top two subgroups), the estimates from these models are strikingly different from the other observational estimates — and close to the experimental estimates.

The variation of the reduction in error for each of the estimates using stratification on propensity scores is further illustrated by Figure 4.6, which displays reductions in error relative to the naive observational estimates. For the less popular half of domains, even the estimates using same domain sharing only achieve approximately 20% error reduction. On the other hand, for the top quartile, we estimate error reduction to be over 80%. We can thus attribute much of the error reduction for the overall analysis displayed in Figure 4.3(b) to error reduction for the most popular quarter of domains.¹⁹

¹⁹Recall that we are estimating treatment effects on the treated, so all estimates are weighted

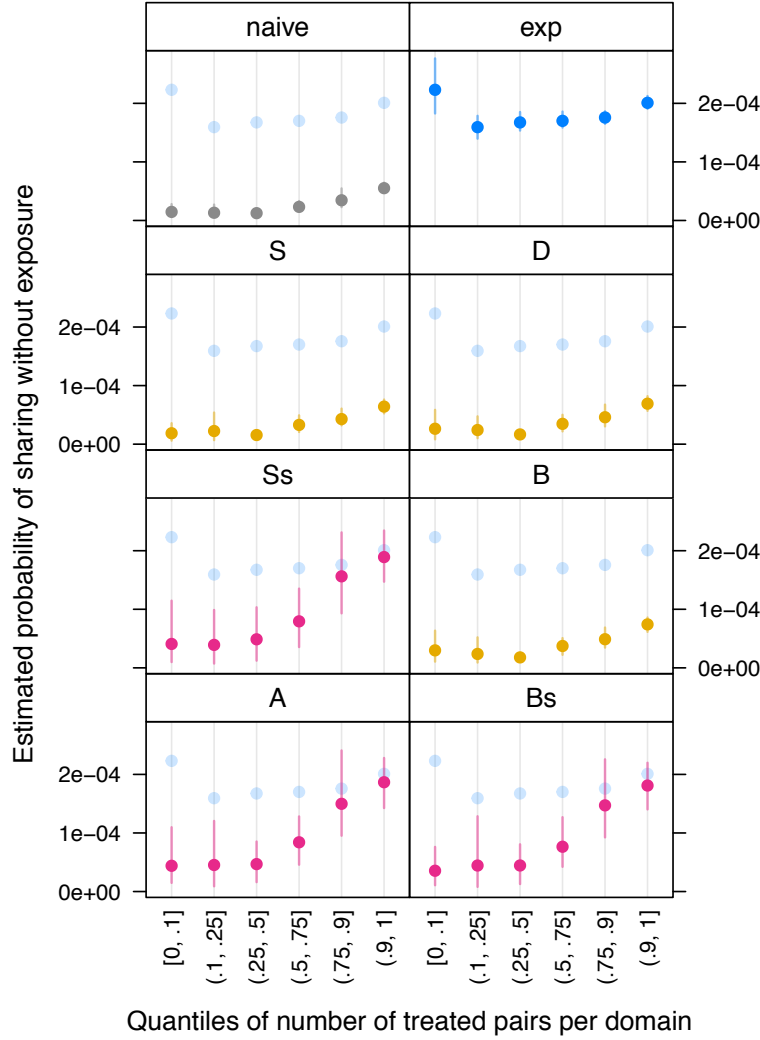


Figure 4.5: Estimated probability of sharing for unexposed individual-URLs pairs, $P(Y = 1 \mid M = 1, \text{do}(M = 0))$, by the popularity of each URL's domain. Experimental estimates are superimposed in light blue for reference. Using a propensity score model with the same domain sharing variable substantially increases this estimated probability only for more popular domains, though it increases the sampling variance in all cases. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.

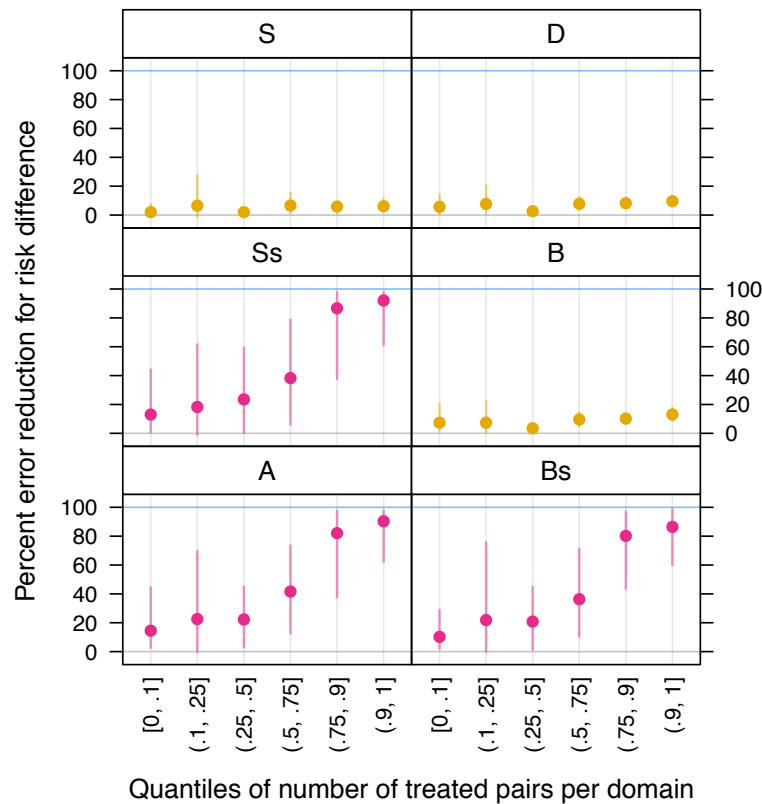


Figure 4.6: Error in estimated peer effects via News Feed using stratification on propensity scores, treating the experimental estimates as the gold standard. (a) Relative error in risk difference estimates for all qualifying URLs. (b) Percent error reduction from the naive observational estimate. Models including same domain shares result in the most error reduction, despite increased variance. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.

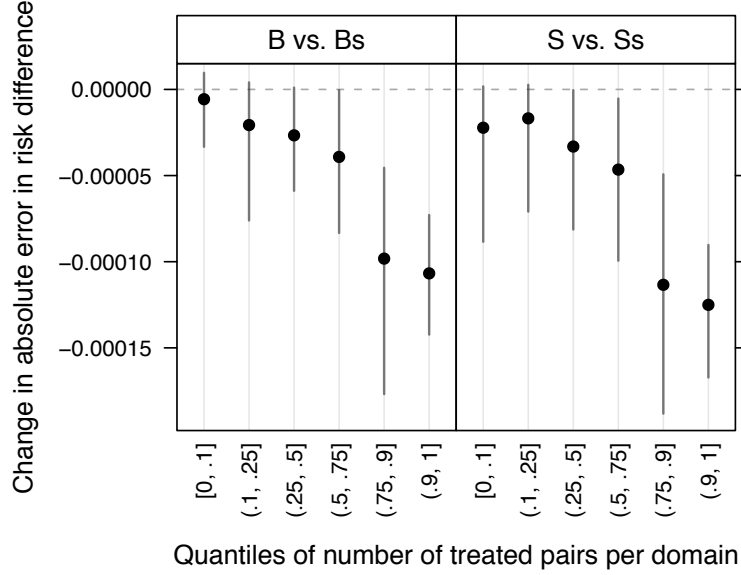


Figure 4.7: Change in error in estimated peer effects from adding same domain shares variable to the B and S propensity score models by the popularity of each URL’s domain. Values less than zero represent reductions in error. For all categories of domains, adding this variable is estimated to reduce error, though this difference is only significant for more popular domains, and it is notably larger for the most popular domains. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.

In the case of the B , Bs , S , and Ss estimates, we can make a direct comparison between estimates that only differ in their incorporation of same domain shares into the propensity score model. Figure 4.7 displays how much this addition reduces error in estimates of ATET. For both models B and S , we estimate that this reduces error for all subgroups of domains, but the reduction of error is both substantial and statistically significant for the most popular quartile of domains.

by the number of exposed individual–URL pairs; thus, the error of the overall estimates is heavily influenced by their performance for the most popular domains. This is a further motivation for the subgroup analysis by domain popularity.

4.3.3 Estimates for individual domains

We now consider the estimates of peer effects via News Feed for individual domains of interest. The subgroup analysis by domain popularity showed how important the most popular domains are for the overall estimates, so this motivates further scrutiny of estimates for these domains. In particular, the following analysis examines our estimates for the 15 domains with the largest number of exposed individual–URL pairs during the experiment.²⁰ We can treat these domains individually because the amount of available data means that we can produce precise peer effect estimates for each, making comparison of experimental and observational estimates possible. They also provide a useful variety of types of URLs that individuals can share.

The probabilities of sharing a URL for individual–URL pairs in the feed and no feed conditions are shown in Figure 4.8. These top domains exhibit substantial heterogeneity in both probabilities. This is, at least in part, attributable to differences in other opportunities for discovering these URLs. For example, URLs at `www.nytimes.com` have a comparatively high probability of being shared, even without exposure via News Feed to a peer sharing the URL; in fact, this proportion is larger than the proportions *with* exposure for several of the other domains. This variation suggests one source for subsequent differences in estimates of peer effects.

We now compare the observational estimates peer effects for each domain (Figure 4.9).²¹ The pattern of estimates for most of the domains is consistent with the

²⁰For this analysis, we exclude two domains that are URL shortening services, as these domains are a very different sort of grouping of URLs than the other domains. The observational methods generally fail to achieve much error reduction for these URLs, when regarding the experimental estimates as the gold standard. However, for reasons related to those discussed in Section 4.3.4, the experimental estimators for peer effects for pairs with these domains probably do not estimate quantities of primary interest.

²¹Additional figures for per-domain analyses are in Appendix A.

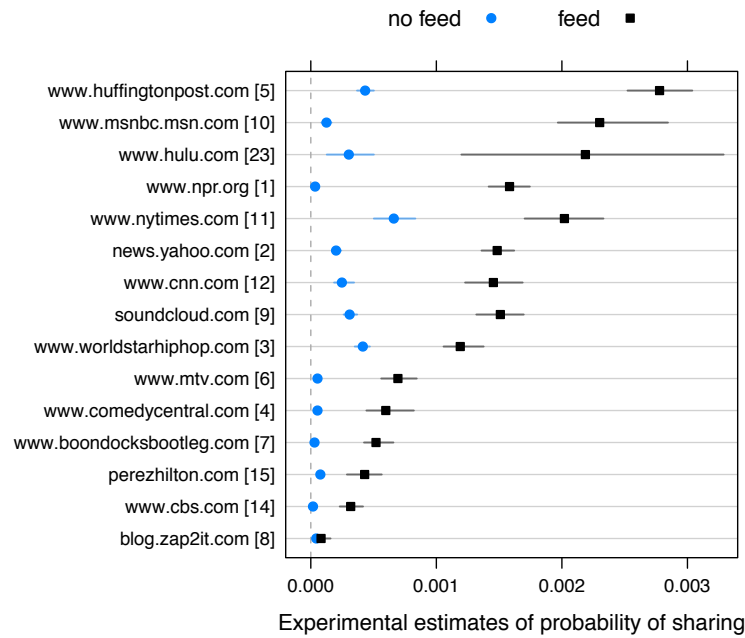


Figure 4.8: Probability of sharing a specific URL in the feed and no feed experimental conditions for the 15 domains with the largest number of exposed individual-URL pairs; numbers in brackets indicate this rank. Domains are sorted by the difference between the two conditions (i.e., the experimental estimate of the risk difference). Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.

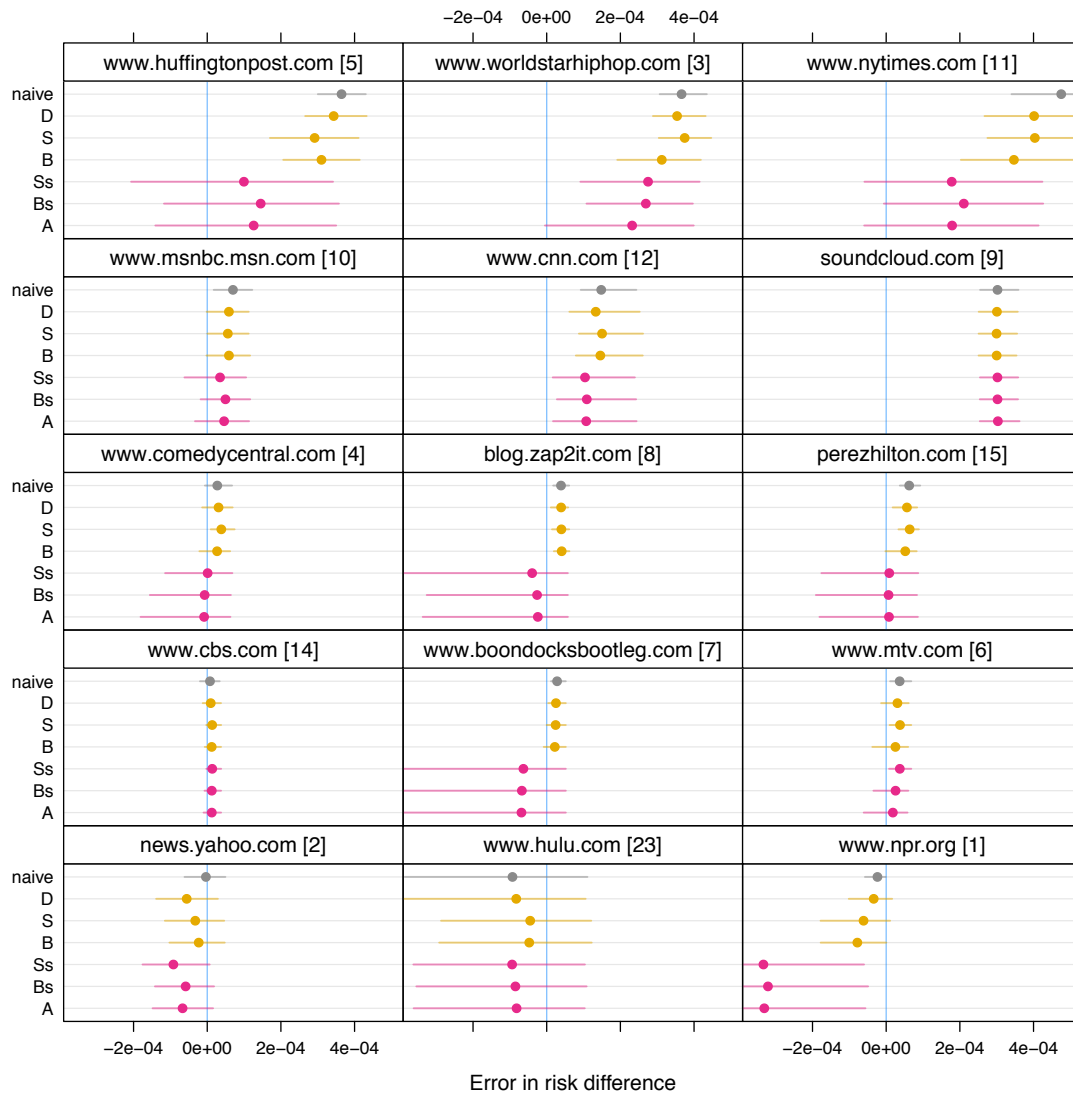


Figure 4.9: Error in estimated peer effects via News Feed using stratification on propensity scores for the 15 domains with the largest number of exposed individual-URL pairs; numbers in brackets indicate this rank. Domains are sorted by largest error. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.

expectation that the observational analyses will produce larger estimates of peer effects. Compared to the experimental estimates, we find that the naive and estimates of the risk difference are significantly larger. On the other hand, there some domains for which the point estimates of this difference suggest that the observational estimates might exhibit *negative* bias. In particular, for `www.npr.org` the naive estimate is marginally smaller than the experimental estimate, and the point estimates from propensity score stratification are also smaller.²² This latter point is consistent with the observed pattern that these models, and propensity score stratification more generally, produce larger estimates of $p^{(0)}$ than the naive estimator; however, it is surprising both that the difference between the experimental and observational estimates has a negative sign and that the propensity score stratification apparently increases this difference.

Since the estimators using same domain sharing have performed well for popular domains so far, we now examine the change in error in risk difference estimates from adding this variable to the propensity score model for each of the most popular domains. Figure 4.10 shows the change in absolute error for each pair of models for each domain. We have colored the points to indicate the sign of the difference between the experimental and naive estimates for that domain. Several domains exhibit the expected pattern of error reduction, and for a few this reduction is statistically significant (e.g., `www.cnn.com`). However, consistent with the comments above, `www.npr.org` shows the opposite consequence: adding same domain sharing

²²This unexpected difference might be a false positive, given that we are conducting several tests here. We do not think this is the case for two reasons. First, the false positive rate is likely lower than 5%. We actually expect not just no difference, but the reverse difference, under which this false positive would be even less likely. Furthermore, the bootstrap confidence intervals we are using are expected to be quite conservative (i.e., having coverage over 95%). Second, as we describe below, there are other reasons to believe that the experimental estimates for `www.npr.org` are, in some important sense, incorrect.

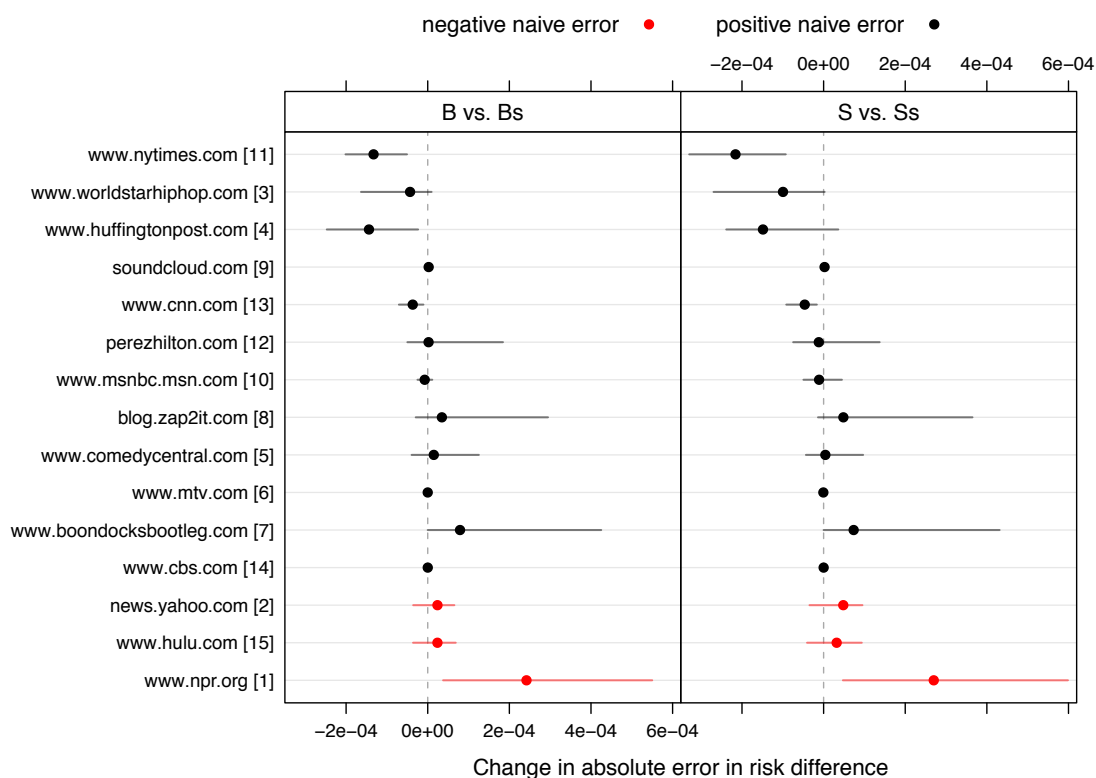


Figure 4.10: Change in error in estimated peer effects from adding same domain shares variable to the B and S propensity score models for the 15 domains with the largest number of exposed individual-URL pairs; numbers in brackets indicate this rank. Values less than zero represent reductions in error. For most of the domains, adding the predictor results in a smaller estimated error. In some cases, this is statistically significant at the level of individual domains. On the other hand, it also substantially increases it for some domains, generally by underestimating peer effects, as can be seen in Figure A.1. Domains are sorted by the naive error in risk difference. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.

to the propensity score model increases error by making the observational estimate of peer effects smaller. We further discuss this and other differences between the observational and experimental results in the next section.

4.3.4 Explaining differences from experimental estimates

For much of the above discussion, we have regarded the experimental estimates as the gold standard. That is, we have regarded the experiment as identifying the average peer effects of interest for those who would be exposed. Thus, discrepancies between the experimental and observational estimates are then attributable to sampling variance in either and bias in the observational estimates. We expected the observational estimates to suffer from confounding bias because of selective tie formation and dissolution (i.e., homophily and heterophily), common external causes, and prior influence. Except for heterophily, these would all make it more likely for peers to share the same URLs, even in the absence of peer effects, so we anticipated that the naive observational analysis would overestimate peer effects, and that the estimators using propensity score stratification would reduce, but not eliminate or reverse, this bias. This is the primary explanation of differences between the experimental and observational estimates. In this section, we consider two other explanations of differences between these estimates.

Total peer effects versus peer effects of exposure for the exposed

Even if total average peer effects are conditionally unconfounded given the covariates used in our propensity score models, the observational and experimental estimates can differ if the former consistently estimate total peer effects (i.e., effects of peer

sharing via all mechanisms) and the latter consistently estimate peer effects of exposure through News Feed and profile feeds. This places an important limitation on what we can learn from this constructed observational study. We nonetheless regard studies such as this as one of the best available tools for better understanding the performance of observational methods for estimating peer effects.

We expect that while exposure via News Feed and profile feeds is not an exhaustive mechanism for peer effects in URL sharing on Facebook, it may be nearly exhaustive, since the other primary mechanism is exposure through that peer sharing the URL on Facebook and then, *because of this prior sharing decision*, sharing with the ego via some other method, such as via email, in person, or through Facebook chat. While sharing via other methods may be common, and this may be associated with sharing on Facebook, we expect that doing so as a result of having also done so on Facebook is relatively rare.

Problems in meaningfully individuating URLs

One additional reason to doubt that the experimental estimates should uniformly be treated as the gold standard concerns how URLs are individuated. While some attempts were made in the original experiment to canonicalize URLs — that is, to identify multiple URLs that correspond to the same online resource or page — this is not perfect for many uses. In particular, there are many cases where there are multiple URLs that point to essentially the same resource. At the extreme, there can be variations in the URL that are directly related to whether a user arrived at it via Facebook. For example, many `www.npr.org` URLs have the form `http://www.npr.org/templates/story/story.php?storyId=[n]` where `[n]` is a numeric identifier for the story. Some of these URLs also appear, or appear exclusively, appended with

the query parameters `&sc=fb&cc=fm`. This variation in the URL does not change the primary content shown to visitors to the URL. Rather, it is apparently used by `ww.npr.org` to track whether visitors are coming to the site from Facebook. For the purpose of studying peer effects in information diffusion, media consumption and sharing, etc., treating these two URLs as distinct, such that an individual is only counted as sharing the same URL if they share a version that matches this appended set of query parameters exactly, is likely undesirable. Consider an individuals who would be exposed to a peer sharing a URL with the query parameters (i.e., $M = 1$). They might encounter the same content through other means, in which case the URL would likely not have these parameters, or have different ones, and share that URL. Under the experimental analysis in Bakshy et al. (2012b) and in this chapter, they would not be counted as sharing the URL. If we would prefer to consider these to be the same URL, then this results in underestimating $p^{(0)}$ and $p^{(1)}$ and likely overestimating their difference and ratio. We regard this as likely a primary cause of the unexpected negative difference between the experimental and observational estimates for `www.npr.org`.

4.4 Discussion

It is often not possible to conduct experiments that identify peer effects. Investigators are then forced to choose between leaving important questions about peer effects unanswered or providing answers based on analysis of observational data. This is the first method evaluation study of observational methods for estimating peer effects that makes use of experimental results as a comparison point. In the current study, we estimated a portion of the average exposure–adoption function for particular subpopulation of interest; specifically, we estimated the average effect of going from

non-exposure to exposure to a single peer adopting the behavior. To our own surprise, we found that some of the observational estimators — in particular, stratification on fitted propensity scores from models including measures of prior behaviors very similar to the focal behavior — reduced bias in average peer effects on the exposed to less than 4% of the experimental estimates and to approximately 20% of the bias of the naive estimates.

On the other hand, our results suggest that the observational estimators performed much more poorly for behaviors for which there was not much data about similar prior behaviors. In particular, treating the experimental estimates as the gold standard, we estimated that the best observational estimators only reduced the bias of the naive estimator by only approximately 20% for peer effects in sharing URLs from less popular domains. Finally, an unexpected negative difference between the experimental and observational estimates led us to identify a problem with individuation of the focal behaviors that affected the credibility of the experimental estimates themselves. This also highlights a direction for future work: since experiments that block mechanisms of peer effects are often expected to *underestimate* total peer effects and observational methods are expected to *overestimate* them, it may sometimes be possible to use them in combination to, under additional assumptions, bound total peer effects.

Together, these results suggest that observational analyses can at least sometimes be quite informative about peer effects. However, analysts who lack measures of prior behaviors closely related to the focal behavior (e.g., same domain sharing) and instead rely on a merely convenient selection of covariates (as in Model *D* in particular) should expect that their estimates of peer effects may suffer from very substantial bias (cf. Shadish et al., 2008). For example, in our overall analysis of peer effects in URL sharing, the experimental analysis concluded that exposure to a peer

sharing a URL made individuals six times more likely to share that URL, while the analysis using only demographic variables concluded that they were exposure made them 24 times more likely to do so.

Chapter 5

Peer effects in the development of cultural rituals

Culture — the totality of socially transmitted preferences, expectations, and practices — emerges from the local activities of individuals embedded in particular social and physical circumstances. Behavioral science has offered some demonstrations of how the clustering of behaviors that is characteristic of culture can emerge from simple dynamical systems (Latane, 1996). In particular, the local homogeneity in behavior which is partially constitutive of cultural rituals can result from individuals' tendency to do as their peers do — that is, from positive peer effects. In order to model the processes by which cultural rituals develop, we investigate peer effects in the use of a communication technology on the Thanksgiving Day holiday in the United States.

For many Americans, Thanksgiving involves ritualized expressions of gratitude for material abundance, relationships with loved ones, major sporting events, and general well-being (Baker, 2009; Wallendorf and Arnould, 1991). Whether and how

people participate in this ritual is expected to be affected by the participation of their peers. In particular, we expect that there will be positive peer effects in adoption of a new medium for expressing gratitude on Thanksgiving Day. Following the economic analysis of Section 2.3, we can hypothesize reasons for preference, expectation, and constraint interactions. Egos may prefer to not be alone in expressing gratitude on Facebook (a preference interaction). Observing peers expressing gratitude may reveal positive consequences of doing so, such as others ‘liking’ the post (an expectation interaction).¹ Finally, egos’ ability to express gratitude on Facebook could be increased by observing examples, which can be easily imitated (a constraint interaction). The present study does not aim to empirically distinguish these different types of interactions, but each provides a reason to expect positive peer effects in this behavior.

Since Thanksgiving is a cultural holiday observed by a large number of people, learning about the causes of changes in its observance over time is of direct value to the study of culture. We may additionally be able to generalize from knowledge about these specific peer effects. First, our results can provide evidence for and against theories that make predictions about peer effects in cultural rituals. Note that the reasons for preference, expectation, and constraint interactions likely apply to other cultural rituals. Second, the magnitude of resulting peer effect estimates can give us reason to expect similarly-sized peer effects in similar behaviors and contexts.

We use a large peer encouragement design to estimate peer effects in the choice to express gratitude on Thanksgiving via Facebook. Thus, this study demonstrates the empirical strategy described in Section 3.4.2.

¹Of course, if individuals received negative comments from peers in response to expressing gratitude on Facebook, this might result in negative peer effects if egos observe these comments and infer that their utility for expressing gratitude might be lower than previously expected.

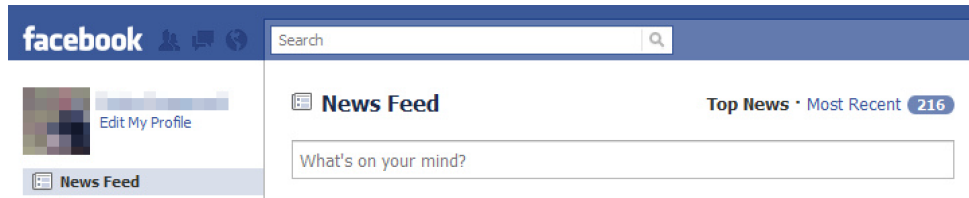


Figure 5.1: Top left portion of the Facebook home page as of fall 2010. The standard (control) prompt is shown. For users assigned to the alternative prompt ($V_i = 1$), the light grey text instead read, “What are you thankful for?”

5.1 Method

In the case of ritual expressions of gratitude via Facebook on Thanksgiving, we expect that characteristics affecting the propensity to engage in this behavior are clustered in the Facebook social network, making it problematic to attribute subsequent clustering in expressions of gratitude to peer effects (Section 3.2). In order to address this confounding, the present work uses a *peer encouragement design* in which some individuals were induced to engage in the focal behavior (i.e., expressing gratitude via Facebook) in order to identify the effect of this behavior on their peers. We also conduct a purely observational analysis for comparison.

We restrict this study to individuals who were both identified as being in the United States and were using Facebook in American English. These restrictions are motivated by November 25, 2010, being observed as Thanksgiving Day primarily by people in the United States and that both encouraging and identifying expressions of gratitude depends on language. Of these individuals, over 70 million used Facebook anytime during 12:00 AM November 25, 2010, through 12:00 AM November 26, 2010, Pacific Standard Time. All of these *qualifying users* are eligible for inclusion as peers in the subsequent analysis.

On the Facebook homepage and elsewhere on Facebook, users can share text

with their friends (a *status update*); they do this by entering text in a box that has the question, in light grey text on a white background, “What’s on your mind?” (Figure 5.1). On Thanksgiving Day 2010, 1.01% of qualifying users were randomly assigned to be presented with an alternative prompt, “What are you thankful for?” The alternative prompt was again replaced with the control prompt if the user posted a status update. This alternative prompt functions as an encouragement to the focal behavior of posting a grateful status update on Facebook.

5.1.1 Expressions of gratitude

Each status update shared by qualifying users was automatically coded as either expressing gratitude or not. The focal behavior of expressing gratitude was operationalized with a regular expression that picks out any uses of ‘thanks’, ‘thankful’, ‘thank’, ‘thanking’, ‘thanked’, ‘grateful’, ‘gratitude’, ‘ appreciate’, ‘appreciating’, ‘appreciated’, ‘appreciative’, or ‘appreciation’.² Note that this regular expression does not count status updates containing only ‘Thanksgiving’ as expressing gratitude; so, e.g., ‘Happy Thanksgiving, everyone!’ is not coded as expressing gratitude. Hereafter, references to grateful or thankful status updates are references to those coded as such.

5.2 Model

Let $V_i = 1$ if and only if individual i is assigned to the alternative prompt, and $V_i = 0$ otherwise. We aggregate these assignments into the total number of peers assigned to the prompt. For most analyses, the relevant set of peers is those Facebook friends

²The regular expression is: `.*\b(thanks|thankful|thank|thanking|thanked|grateful|gratitude|appreciate|appreciating|appreciated|appreciative|appreciation)\b.*`

who logged into Facebook before the ego, as all other Facebook friends cannot have been affected by the prompt by the time the ego logs in. Furthermore, preliminary analysis suggested that peers who only log in very shortly before the ego are not induced to post grateful status updates by the time the ego logs in. Thus, we have that \mathbf{A} is the adjacency matrix for the directed friendship graph such that $A_{ij} = 1$ if and only if individuals i and j are Facebook friends on November 24, 2010, and i logs in on Thanksgiving at least an hour after j . Hereafter, an ego's *peers* are only those Facebook friends who log in an hour before the ego.³ Thus, the number of encouraged peers is

$$Z \equiv \mathbf{A}\mathbf{V}.$$

This is our instrumental variable, which we use to test for and identify peer effects in expressions of gratitude. The number of prompted peers Z is only ignorable conditional on the total number of peers who log in before the ego

$$F \equiv \mathbf{A}\mathbf{1}_N,$$

which is the ego's out degree in the graph defined by \mathbf{A} .

Let $Y_{i,(t,t+1]} = 1$ if and only if individual i posts a thankful status update during the period $(t, t + 1]$, and $Y_{i,(t,t+1]} = 0$ otherwise. Let T_i be the first time (in hours) individual i logs in during Thanksgiving Day. We treat whether an individual posts a thankful status update in the hour after first using Facebook on Thanksgiving Day as the primary outcome; that is, we are concerned primarily with the binary outcome

$$Y_i \equiv Y_{i,(T_i, T_i+1]}.$$

³We selected this time threshold for determine the set of peers in part based on errors in measurement of login times (see footnote 4) and in part because of the observed timing of the shock of the prompt to the prompted users (Figure 5.4).

We only estimate peer effects for egos who did not use Facebook between 9:00 PM November 24, 2010, and 12:00 AM November 25, 2010, since this includes many users who used Facebook during Thanksgiving Day, but could have any peers who were prompted.

Let \mathbf{X} be a matrix of covariates, including at minimum a column of 1s. Some analyses include gender, as specified on the individual's profile or otherwise inferred, and the number of status updates posted by an individual in the prior month as covariates. We specify which covariates are used for a particular analysis with the results.

We are interested in summarizing the causal relationship between peer behavior D and ego behavior Y . Written as a nonparametric structural equation, this is

$$Y_i \leftarrow f_Y(D_i, F_i, X_i, U_i)$$

where \mathbf{U} are all the unobserved variables that affect Y . Note that Z_i is excluded from the structural equation for Y_i . The structural equation for D is

$$D_i \leftarrow f_D(Z_i, F_i, X_i, U_i).$$

We use the exclusion restriction encoded in these structural equations to test for peer effects in expressions of gratitude. The causal relationships among all these variables are represented in Figure 5.2.

As reviewed in Section 3.1.5, the assumptions encoded in these structural equations and Figure 5.2 are insufficient to identify, rather than only test for, average peer effects. Either of two additional assumptions, commonly used in instrumental variables (IV) analysis but which we now specify for this case, are necessary for this

purpose. First, we can assume that the peer effects are homogenous; that is, we can assume that a unit increase in D causes the same increase in the probability of the ego posting a grateful status update for all egos, at least conditional on additional variables.

Peer effect homogeneity. $(d - d')\Delta = E[Y_i \mid \text{do}(D = d)] - E[Y_i \mid \text{do}(D = d')]$ for all $d, d' \in \{0, 1, 2, \dots\}$

This is a restrictive assumption in that it requires that the peer effects be the same for all egos within a subgroup and be linear on some particular scale (as written here, the probability scale). Thus, it is desirable to consider adopting the alternative assumption of monotonicity:

Prompt monotonicity. In terms of structural equations for D , $f_D(z', F_i, X_i, U_i) \geq f_D(z, F_i, X_i, U_i)$ for all i and $z, z' \in \{0, 1, \dots, Z_i\}$ such that $z' > z$.

Monotonicity is more often stated in terms of potential outcomes as follows. Where $D_i^{(z)}$ is the value of D_i that would be observed with the instrument Z set to z , then monotonicity requires that, with probability 1, $D_i^{(z')} \geq D_i^{(z)}$. In this context, this is the assumption that prompting additional peers can only increase the number of peers that post a grateful status update.

5.2.1 Estimation and dependence

Some aspects of this model introduce non-standard estimation problems. In particular, Z is only ignorable conditional on F . Thus, one approach to estimating effects of Z and using Z as an instrument is to nonparametrically condition on each of the values of F . This is the approach we take throughout. This motivates us to exclude

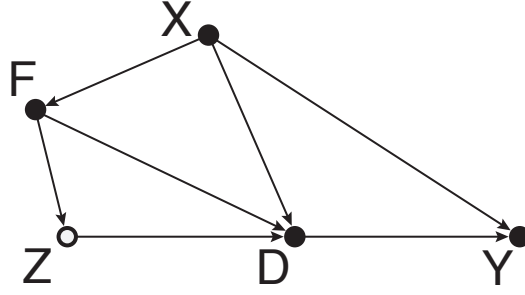


Figure 5.2: Causal DAG representing the relationships among variables in the Thanksgiving Day study. All of the solid black nodes are expected to have common causes U , which are not shown to make the diagram considerably simpler. On the other hand, the instrument Z is colored differently because its only cause, F , is shown. Z is a valid instrument when conditioning on F . The exclusion restriction is encoded by the lack of a path from Z to Y except via D .

from analysis egos with values of F in the top decile (over 215) because the data is very sparse here.

The other approach to making Z ignorable is to account for the probability of observing each individual i with each value of Z . Each Z_i is the sum of F_i Bernoulli trials with $p \approx 0.01$, so to the extent that p is known exactly, we can use these probabilities in Horvitz–Thompson estimation (Aronow and Samii, 2011).

Note that while the distribution of each Z_i is binomial, the marginal distribution of Z is overdispersed since the Z_i are not independent: two individuals i and j who have peers in common have that same number of Bernoulli trials in common. Given how we have defined the peer relationship to incorporate individual-level visitation behavior, there is less of this dependence than in the full friendship graph. Aronow and Samii (2011) provide conservative (i.e., positively biased) variance estimators for designs in which assignment to different exposures is not independent. These

estimators are difficult to use, as they require computing probabilities of mutual exposures for all pairs of units, and our design does not feature the substantial dependences they were designed for. Lacking reasonable alternatives, we produce confidence intervals using methods that ignore this dependence.

5.3 Results

The structure of our analysis is as follows. We first examine the effect of the alternative prompt on the individual users who are prompted, where we find that this prompt causes expressions of gratitude and causes them to occur earlier. This facilitates understanding how the encouragement works. This analysis carries over to ego's peers considered in aggregate, where an increase in prompted peers causes an increase in peer expressions of gratitude. Then by treating D_i as formalized above as the exhaustive mechanism by which prompting peers affects ego expressions of gratitude, we estimate peers effects. This instrumental variables analysis allows estimation of peer effects in expressions of gratitude.

5.3.1 Prevalence and effect of prompting individual users

Assignment to the alternative prompt causes users to post grateful status updates and to post them sooner in the day. Of qualifying users assigned to the control prompt, 8.9% post a grateful status update after logging in (i.e., they have $Y_{i,(T_i, \cdot]} = 1$) and 4.8% do so within the first hour after logging in ($Y_{i,(T_i, T_i+1]} = 1$). Of users assigned to the alternative prompt, 10.8% post a grateful status update after logging in and 5.9% do so within the first hour after logging in.

Ritualized expressions of gratitude via Facebook are thus neither very rare nor

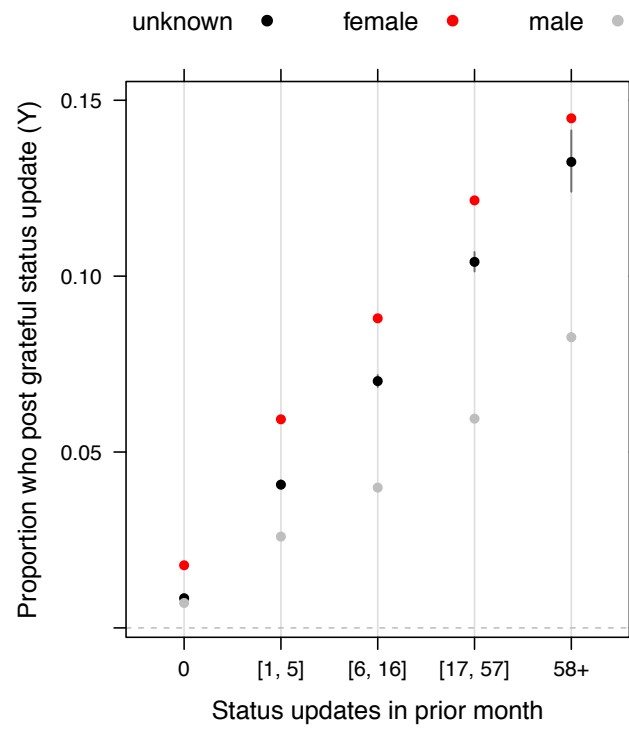


Figure 5.3: Probability of ego behavior as a function of observed ego characteristics — gender and previous posting of status updates. Being female and previously posting status updates both predict posting a grateful status update within the first hour after login. Error bars are 95% confidence intervals.

widespread among qualifying users. As might be expected, posting a grateful status update is much more common among individuals who have recently posted status updates. Figure 5.3 shows how this behavior varies with use of this medium in the prior month and gender.

The effect on the commutative proportion of users who post a grateful status update is presented in Figure 5.4, which compares the empirical CDFs of the time since login at which users assigned to the two prompts post their first grateful update. Here we can see that most of the effect occurs shortly after a user's first login on Thanksgiving.⁴

These results aid understanding of how the prompt functions as an encouragement to the focal behavior: it primarily serves to (a) cause individuals to post a grateful status update shortly after login who would not have and (b) cause other individuals to post a grateful status update earlier than they would have otherwise. Both of these effects are expected to result in exogenous variation in the number of peers who post grateful status updates before an ego's first login.

5.3.2 Aggregate effect of prompting peers on peers

It is the number of encouraged peers, Z_i , and peers who post grateful status updates before ego i logs in, D_i , that enter into our model. Thus, we now turn to characterizing the aggregate effect of prompting peers on prior expressions of gratitude by peers.⁵

⁴ Both ECDFs are not exactly zero before the recorded login time. This indicates that first login times are measured with error. This may happen due to third-party applications used to interact with Facebook. We cannot remove individuals with a status updates before their first login time, since this would be conditioning on a variable very similar to our established outcome.

⁵In the context of IV analysis, this analysis is described as constituting the *first stage*, in which the effect of the instruments on the treatment is examined.

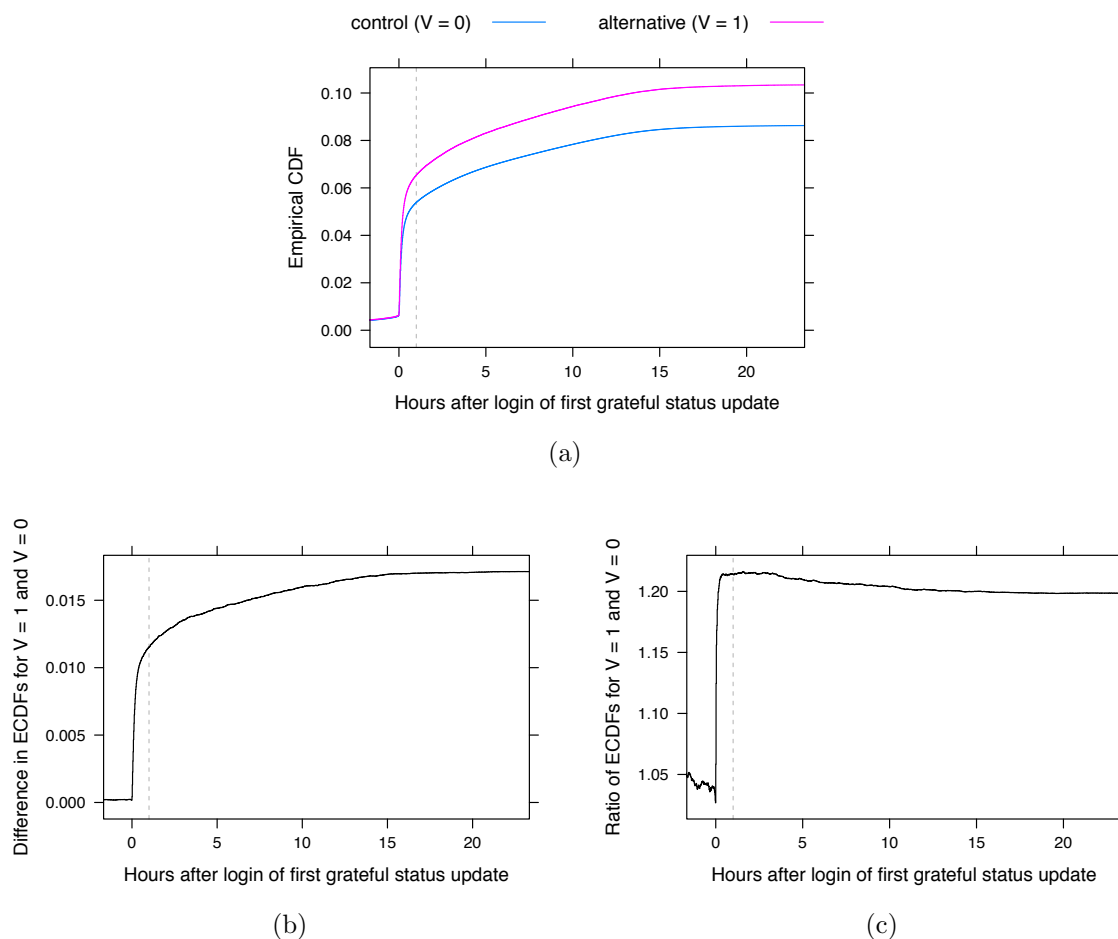


Figure 5.4: Effect of prompting individual users on occurrence and timing of posting grateful status updates. (a) Cumulative proportion of individuals with the control and alternative prompt posting at least one grateful status update by a number of hours after login. (b) Probability of alternative prompt causing first grateful post to occur prior to some number of hours after login. (c) Ratio of cumulative proportions. Much of the difference in adoption of the behavior between the control and alternative occurs shortly after login, and during the first hour after login (dashed line) in particular. The ECDFs are not exactly zero prior to the login time due to errors in measurement of first login times.

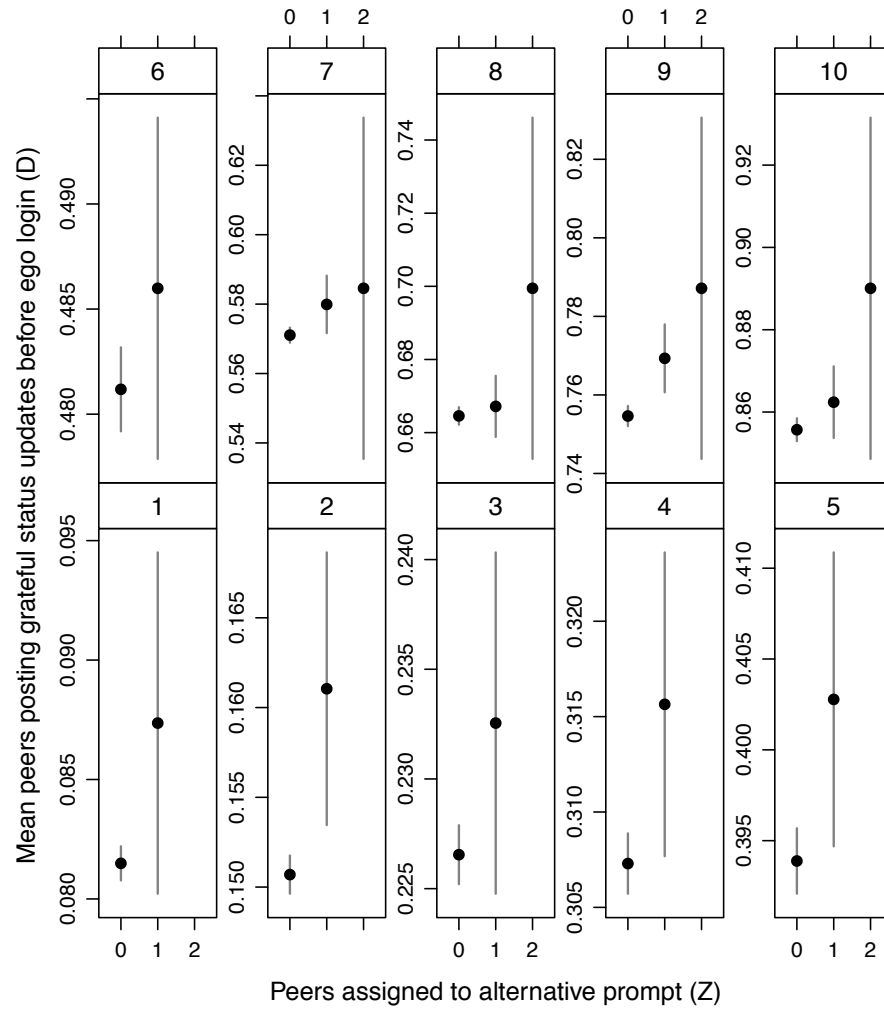


Figure 5.5: Aggregate peers expressing gratitude, D , by number of prompted peers, Z , and number of peers, F , for the first decile of F . Error bars are 95% confidence intervals and implicitly indicate variation in the number of egos for each value of Z . Some combinations of Z and F with very little data are excluded.

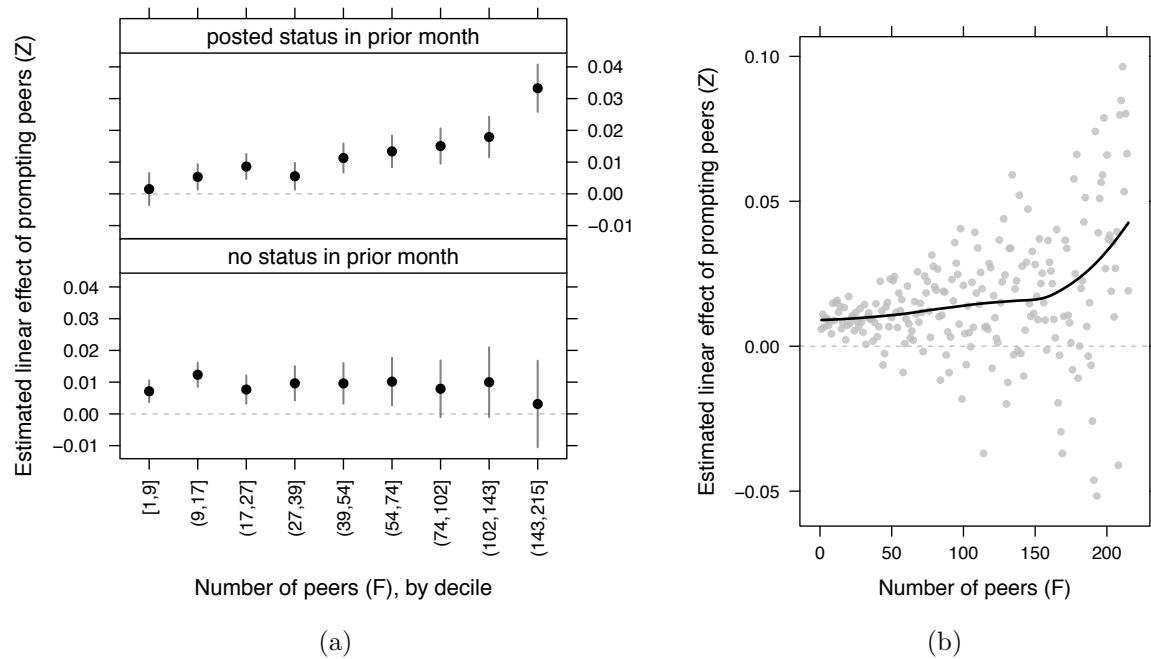


Figure 5.6: Average linear effect of prompting peers, Z , on peer expressions of gratitude, D . (a) Estimates aggregated by deciles of F , excluding the top decile, and whether the ego has recently posted a status update. The model for each subgroup includes fixed effects of each value of number of peers, F . The effect of prompting an additional peer varies with F . Error bars are 95% confidence intervals. (b) Estimates for each value of F summarized by a local regression weighted by the number of observations.

Figure 5.6 presents estimated linear effect of Z on D for each number of peers F for which we have sufficient data. Consistent with the analysis of the prior section, we find generally positive effects of prompting peers on peer expressions of gratitude, such that prompting an additional peer causes 0.0081 to 0.023 additional peers to post a grateful status update before the go logs in. This provides further evidence that the encouragement is working as expected and is useful as an instrumental variable to test for and estimate peer effects.

5.3.3 Effects of peer behaviors on egos

Having examined how the encouragement produces exogenous variation in peer behavior, we now use instrumental variable estimators to estimate peer effects using this variation. In particular, we estimate a linear approximation to the average exposure–adoption function for egos whose peers are induced by the prompt to post a thankful status update.

Technical description of estimators

The most straightforward estimator we applicable here is two-stage least squares (TSLS).⁶ Conceptually, this can be understood as fitting a OLS model for D :

$$D = V\delta + Z\gamma + \eta,$$

⁶Given that Y is binary, some analysts would prefer to use a model for binary outcomes to avoid predictions outside the unit interval and to increase asymptotic efficiency. We follow Angrist (2001) and others in regarding average effects, rather than index coefficients, as being of primary interest and preferring the interpretability of a linear probability model.

storing the fitted values \hat{D} , and entering these, along with covariates, into a OLS model for y :

$$Y = \mathbf{V}\alpha + \hat{\mathbf{D}}\beta + \varepsilon.$$

The TSLS estimator for β can be written as

$$\hat{\beta}_{TSLS} = (\tilde{\mathbf{D}}^T \mathbf{P}_z \tilde{\mathbf{D}})^{-1} (\tilde{\mathbf{D}}^T \mathbf{P}_z \tilde{Y})$$

where $\tilde{\mathbf{D}}$ and \tilde{Y} are versions of \mathbf{D} and Y with the covariates \mathbf{V} partialled out and $\tilde{\mathbf{P}}_z = \tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}^T \tilde{\mathbf{Z}})^{-1} \tilde{\mathbf{Z}}^T$; that is, \mathbf{P}_z is the “hat” matrix that projects onto the columns of $\tilde{\mathbf{Z}}$, where $\tilde{\mathbf{Z}}$ is \mathbf{Z} with the covariates \mathbf{V} partialled out. We report confidence intervals using a heteroskedasticity-robust “sandwich” estimate of the covariance matrix.

IV estimates of peer effects

Our primary IV analysis produces estimates for each combination of deciles of F and whether an ego has posted any status updates in the month prior to the experiment; that is, \mathbf{X} is made up of an intercept and indicators for all these combinations and the instruments are Z interacting with these indicators, $\mathbf{Z} = \mathbf{X} \otimes Z$. We include indicators for the values of F , such that the matrix of covariates is $\mathbf{V} = [\mathbf{X}\mathbf{F}]$, where \mathbf{F} is a matrix of indicators for values of F .⁷

The resulting estimates are presented in Figure 5.7. Only two of the estimates are significantly different from zero.⁸ Egos in the first decile of F who had not posted a status update in the prior month have an estimate of 0.019 ($z = 4.8$, $p < 0.0001$, 95%

⁷Of course, one value of F per decile is not included, so that \mathbf{V} remains full rank.

⁸Since there are 18 tests in this table, one might worry about inflated error rates from multiple testing. Note that both significant estimates are highly so and would remain significant even after a conservative correction.

CI [0.011, 0.027]). That is, an additional peer posting a grateful status update causes a 1.9% absolute increase in their probability of doing so also. The other statistically significant estimate is for egos at the other extreme: egos with a large value of F who have recently posted a status update.⁹ We estimate a much smaller peer effects for these individuals, as might be expected from the idea that a single peer posting a grateful status update might be less significant or even not noticed when many other Facebook friends have already been active earlier that day.

In interpreting these estimates, we should recall that if peer effects are heterogeneous within these subgroups, as they likely are, then these can only be expected to consistently estimate a local average treatment effect (LATE) for egos whose peers are induced by the prompt to post grateful status updates before the ego logs in.

How should we understand the substantial variation in the precision of these estimates? Some of this variation can be attributed to differences in sample size because of association between peer count and prior behavior. It is also due to differences in the effects of prompting peers on peer behavior: as described in the previous section, egos with more peers have a stronger “first stage” and egos who post status updates have friends who are more likely to post status updates and be more responsive to the prompt (see Figure 5.6). This translates into increased precision in these estimates.

5.3.4 Observational analysis

We can also conduct a purely observational analysis of the association between peer and ego expressions of gratitude. This can be regarded as a, possibly naive, estimate of peer effects, which we compare with our estimates using the peer encouragement

⁹Results for an additional model specification are included in Appendix B. Estimates for the two subgroups highlighted here are similar.

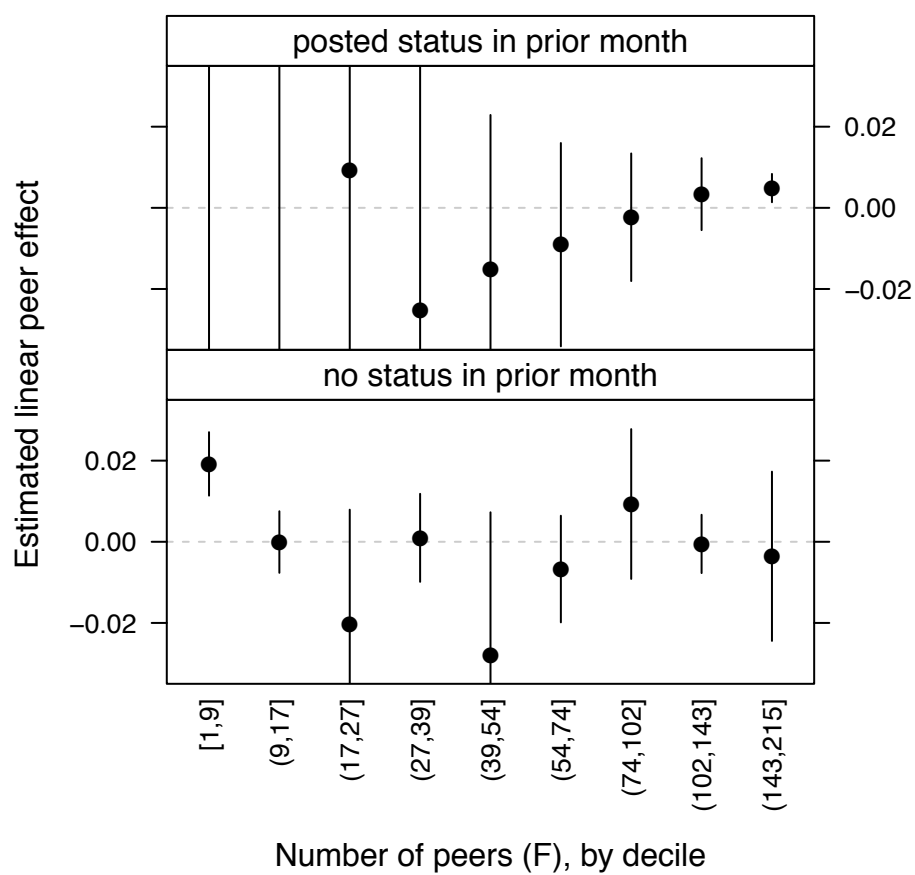


Figure 5.7: Instrumental variable estimates of peer effects by deciles of number of peers, F , and prior posting of status updates within the first hour after login. Estimates for some of the subgroups are relatively imprecise both because of varied sample sizes and differences in how much variation in D is caused by Z within that subgroup. Error bars are 95% confidence intervals.

design. Because we have relatively detailed data about peer and ego behavior (e.g., D is the number of peers posting grateful status updates before the ego logs in), this observational analysis may be able to avoid problems of reciprocal causation that could otherwise result in biased estimates. Furthermore, our use of the encouragement design already required conditioning on F to make Z ignorable, and we included other covariates in our analysis; thus, it may be that conditioning on these variables will reduce confounding bias enough so as to produce informative observational estimates.

Figure 5.8 shows how the probability of the ego positing a grateful status update in their first hour after login varies with peer behavior. These are simply the unadjusted sample proportions. This probability increases substantially for small values of D . These increases are expected to reflect association between the number of peers F and ego behavior. We can additionally condition on the number of peers F , as we did in analyses of previous sections. Figure 5.9 shows these proportions for each value of F in the first decile.¹⁰ The observed increases in the probability of the ego behavior are smaller than in the unconditional analysis.

We can additionally use ordinary linear regression (OLS) to produce observational estimates of linear peer effects. As with the IV analysis, we allow the peer effect to vary by deciles of F and prior status posting. We also include centered indicators of F interacting with D so as to condition on F while estimating an unweighted conditional expectation function, rather than a precision-weighted one. These estimates are shown as a function of number of peers and prior status posting behavior in Figure 5.10. The causal interpretation of these estimates is that an additional peer positing a grateful status updates results in a 1.7% absolute increase in the probability of the ego doing so for those subgroups most responsive to peer behavior.

¹⁰Additional related figures appear in Appendix B.

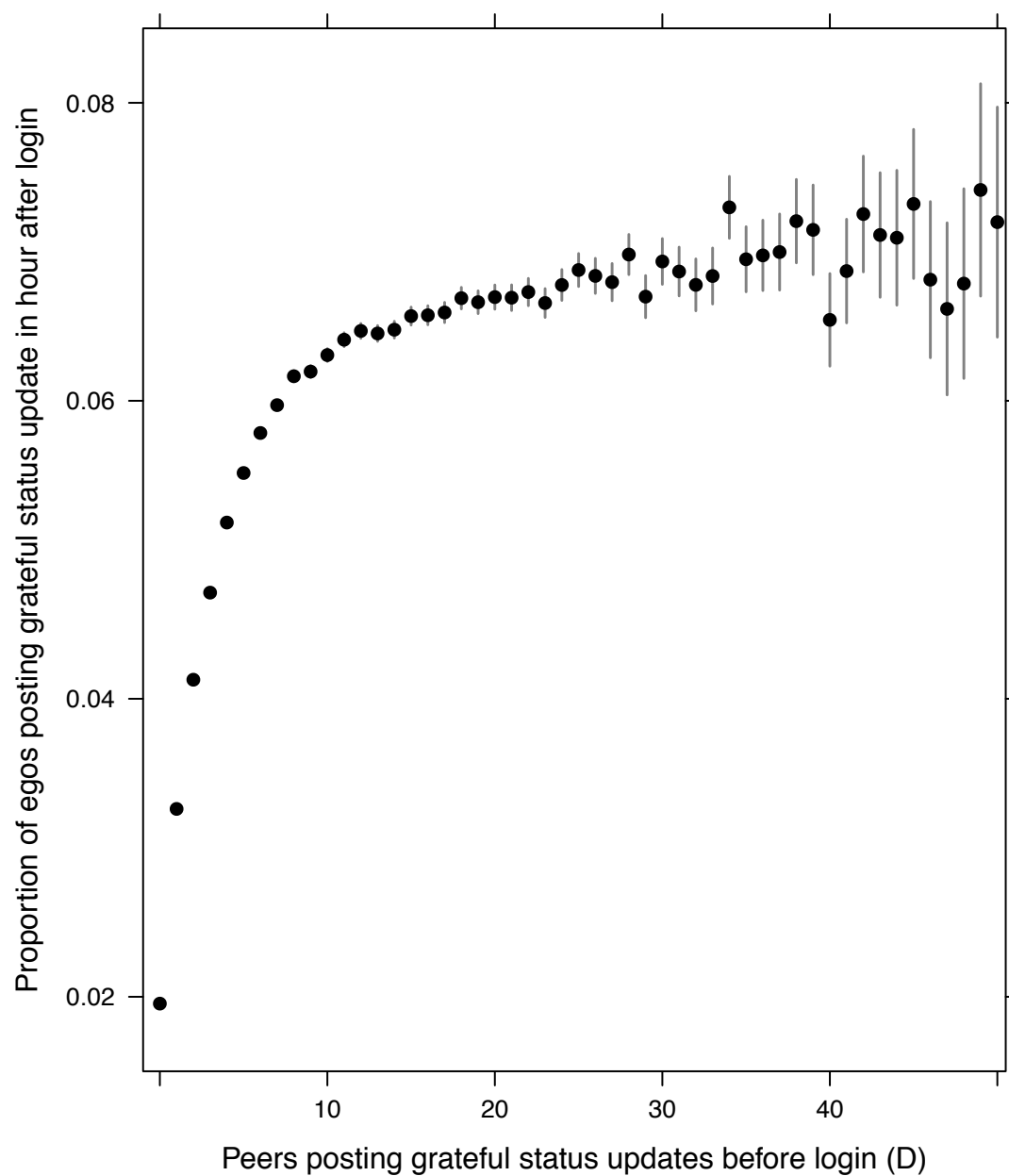


Figure 5.8: Proportion of egos posting a grateful status in the first hour after login as a function of observed number of peers posting a grateful status update before the ego's login, D . Error bars are 95% confidence intervals.

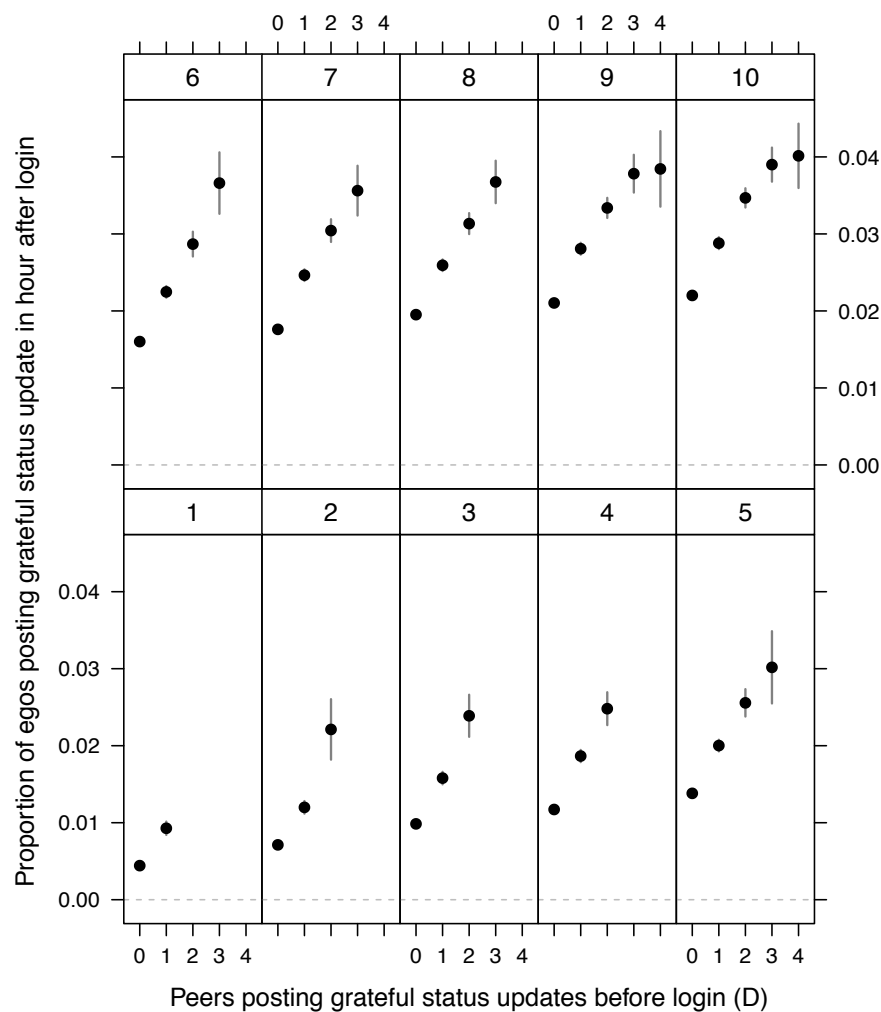


Figure 5.9: Proportion of egos posting a grateful status in the first hour after login as a function of observed number of peers posting a grateful status update before the ego's login, D . Each panel is a number of peers who log in an hour before the ego, F , with $F \leq 10$. Error bars are 95% confidence intervals.

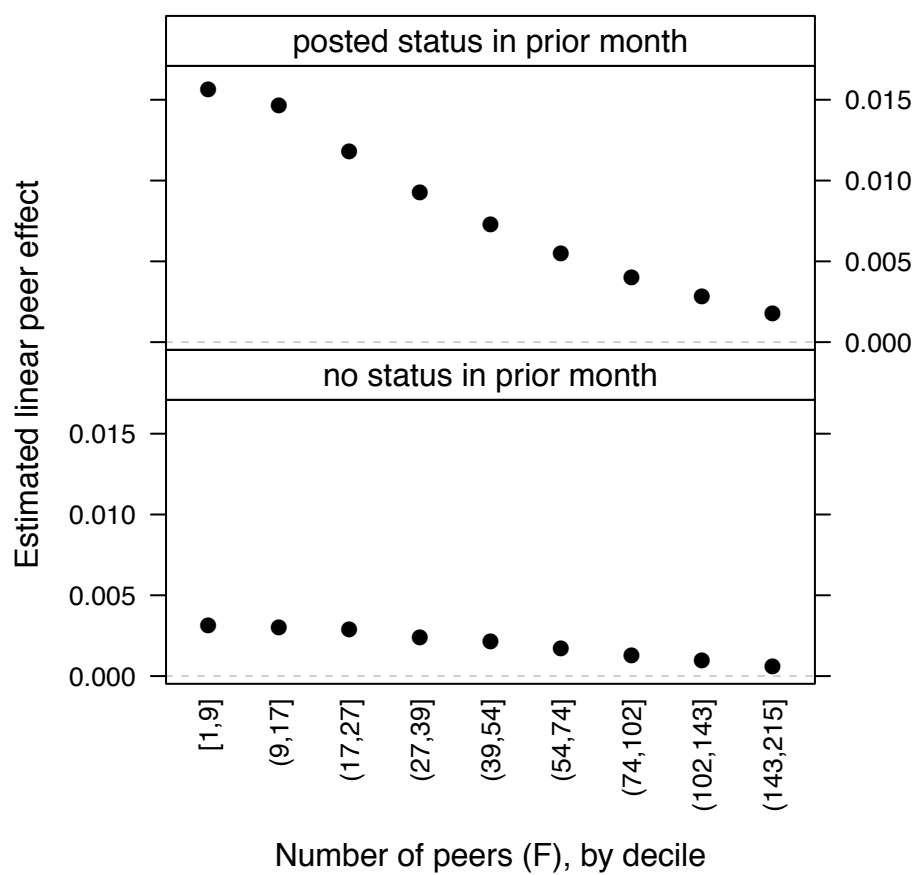


Figure 5.10: Estimated peer effects from a purely observational analysis by deciles of number of peers, F , and prior posting of status updates. This is a OLS regression of Y on D and indicators for each value of F . The indicators for F are centered and interacted with D so as to, under ignorability, estimate the ATE rather than a precision-weighted average. Error bars are 95% confidence intervals.

The estimated effects are much smaller for egos who have not posted status updates in the prior month. As we saw previously, these individuals are also less likely to post a grateful status update in general.

Comparison with IV estimates

In most cases, these observational estimates are consistent with the those from the previous analysis, but this is largely attributable to the imprecision of the IV estimates. On the other hand, the most precise IV estimates are substantially different from the OLS estimates.

How should we interpret these differences? If we believed that peer effects were homogeneous within these subgroups, then we could conclude that, if the instrument is valid, the observation analysis suffers from bias. Interestingly, this bias would be negative for both of the subgroups for which the IV estimates are significant. The IV estimate is six times the OLS estimate for egos with $F < 10$ who haven't recently posted a status update and 2.7 times the OLS estimate for egos in the ninth decile who have recently posted a status update. The direction of difference is more difficult to explain as due to confounding bias from, e.g., latent homophily.¹¹

However, since it is more likely that there is substantial heterogeneity in peer effects, even within the subgroups, there are multiple ways to explain this difference. First, there may still be substantial confounding of peer and ego behavior that is biasing the observational estimate. Second, the LATE that is the estimand of the IV analysis may be for a subpopulation that is different than the whole population over which the observational regression is averaging. Perhaps individuals who are

¹¹We might worry that by only examining the statistically significant IV estimates, we are conditioning on having a large estimate, and that this explains why the IV estimates are larger. However, note that these are not simply larger than the IV estimates for other subgroups, but they are also more precise.

responsive to the prompt are peers of egos who are more susceptible to social influence. This is one plausible explanation of how the IV estimates are larger than OLS estimates.

5.4 Discussion

Our results are consistent with the presence of positive, if modest, peer effects in ritual expressions of gratitude on Facebook on Thanksgiving Day. Both the IV and observational estimates support the conclusion that an additional peer posting a grateful status update increases the probability of the ego doing the same; the largest estimates of this peers effect are just under 2% for both analyses.

The IV estimates suffer from large sampling variance. This both limits our ability to reach precise conclusions from them about the magnitude of peer effects and to make sharp comparisons with the observational estimates. However, there were some substantial differences. Somewhat surprisingly, the two significant IV estimates were larger than the OLS estimates. One explanation of this difference is that the IV estimates are for a subpopulation of egos with peers who are susceptible to the prompt, and that these egos are also more susceptible to peer influence in communication behavior on Facebook. Thus, the peer effects in this group could be much larger than the population average.

We regard the IV estimates as informative about peer effects for a population of substantial interest. One way of applying estimates of peer effects is to use them to plan to effect adoption among some individuals in part by counting on them being influenced by peers who are induced to adopt. Thus, the peer effects that can be used to drive adoption of the behavior or produce are precisely effects for egos with peers who can be induced to adopt. To the extent that planned inducements are similar

to the encouragement used in this research design, we expect the IV estimates to be very informative.

We have intended to both illustrate the opportunities presented by peer encouragement designs and to detail some of the issues in actually using them. Peer encouragement designs have a simplicity — randomly encourage some individuals to adopt a behavior and see how their adoption affects others. We hope we have highlighted the potential of these designs for testing for and estimating peer effects generally and in online communication behaviors more specifically.

Thanksgiving is observed by a large population, so knowledge about peer effects in changes in its observance over time is directly valuable to scientific efforts to understand culture. However, we may also wish to generalize from these results to peer effects in other behaviors. This requires theory that suggests which behaviors and situations are importantly similar to expressions of gratitude on Facebook on Thanksgiving Day. Unfortunately, available theory suggests many possible characteristics of behaviors and situations that may matter. Furthermore, we do not directly observe many of these characteristics. For example, the observed peer effects may have been substantial driven by either preference or constraint interactions, but we do not observe which. Another situation may lack substantial constraint interactions if all individuals are already familiar with the behavior and able to perform it easily. Thus, we should be cautious in reaching broader conclusions about the magnitude of peer effects in other behaviors from this single study.

5.4.1 Limitations

A primary limitation of the IV analysis of this peer encouragement design is the lack of precision in our estimates of peer effects. Despite a large number of observations,

we were left with wide confidence intervals around the estimates of peer effects for many subgroups.

Violations of the exclusion restriction

Peer encouragement designs make the exclusion restriction plausible because the encouragement is applied to different individuals than the outcome is measured on, the encouragement likely only affects the egos through peer behavior. In our estimation of peer effects, we assumed a more specific exclusion restriction that may be violated for multiple reasons. In particular, we assumed that Z only affects Y through D .

We coded status updates as expressing gratitude using a very simple rule. This has the advantage of being completely automated, privacy preserving, and readily interpretable, but it likely also introduced some systematic error into our measurement of gratitude. One of the consequences of measurement error of the treatment is violation of the exclusion restriction: some peers' grateful status updates were not coded as such, but some were nonetheless a consequence of the peer encouragement and likely a cause of ego behavior. That is, this measurement error can also result in a violation of the exclusion restriction.

Since peer behavior occurs over time, there are also likely violations of the exclusion restriction from peers who log in before the ego but are induced to the behavior *after* the ego logs in. This motivated our selection of the peer set for dog i to be determined an hour before the ego logs in $T_i - 1$. Similarly, this was one motivation for examining ego behavior only shortly after their login, $Y_{i,(T_i,T_i+1]}$.

Chapter 6

General discussion

Peer effects figure centrally in both the theories and applications of the social sciences. Until recently, they have been difficult to study *in situ*. This dissertation has contributed to the study of peer effects by developing, evaluating, and applying methods for identifying peer effects by making use of new data sources, opportunities for experimentation, and tools for reasoning about causality. In particular, this dissertation has relied on at least two important ongoing trends in science: large-scale experiments with and observations of human communication behavior, and tools (e.g., formalism, theorems) for causal inference. We expect that future work on peer effects will continue to rely on and advance these trends. In this section, we offer some more general discussion of how the present work fits into trends in computational social science and causal inference.

6.1 Causal inference

The social, behavioral, and economic sciences have long been concerned with reaching causal conclusions — after all, the theories being developed and tested in these fields are causal. But the methods by which this has historically been done have been largely informal. Pearl (2009b) even attributes much current confusion about causal inference in social science to the lack of an alternative to the equals sign ‘=’ for use in equations regarded to be causal (i.e., structural equations) — that is, to the absence of novel formalism for expressing causal claims.

In recent years, however, two developments have changed this situation to some degree. This dissertation relies on and intends to advance these developments. First, econometrics has seen increasing concern with identification of causal quantities through field experiments and natural experiments, which can make required conditional ignorability assumptions more plausible (Angrist and Pischke, 2010). Not only did this result in econometricians and statisticians proving results about, e.g., instrumental variables, as used in our work with peer encouragement designs, but this has increasingly led to more widespread focus on making causal inference in social science more credible (e.g., Morgan and Winship, 2007). We attribute at least some of the recent interest in method evaluation studies, as discussed in Chapter 4, to the additional scrutiny given to conditional ignorability assumptions in empirical work.

Second, there is increasingly cross-disciplinary communication about causality, causal discovery, and causal inference. The work of Judea Pearl (e.g., Pearl, 2009b) has introduced computer scientists and others to these questions and described connections between multiple frameworks for reasoning about causality. For example, in Chapter 3 and the empirical chapters of this dissertation, we used the potential

outcomes, do-calculus, and non-parametric structural equation approaches as each was fitting. Especially in the absence of a perfect randomized experiment, we regard these a critical tools for reasoning about what proposed research designs can teach us about causal quantities.

6.2 Computational social science

Many new technologies, especially communication technologies, are producing data about human behavior at new levels of detail and scope. Many investigators have recently identified the use of these new data sources as constituting the beginnings of “computational social science” (e.g., Lazer et al., 2009). We also highlight that these technologies have not simply made observation easier, but they have also enabled conducting large experiments that affect communication behavior; given the increasing focus on credible causal inference, we regard this as a critical additional feature of this proposed interdisciplinary field.

Lazer et al. (2009) suggest that the most difficult challenges in computational social science may be ethical — and are related to privacy in particular. We would add that since credible computational social science will generally require experimentation, it also raises familiar questions about the ethics of conducting experiments that can affect human well-being.

Increased internal validity (from use of experiments and formal causal inference) and increased external validity (from use of massive real-world data sets drawn describing populations of widespread interest) may put social scientists in the position of producing results of increasing influence on decisions by corporations and governments interested in predicting and controlling the behavior of citizens, consumers, and others. This suggests that investigators should be clear about the value and

costs of likely applications of their work in computational social science. That is, these two trends — in causal inference and computational social science — may shift additional credibility, attention, and responsibility onto social scientists. In particular, whether in public health, marketing, or politics, information about peer effects is of substantial value to decision-makers; we hope that scientists studying peer effects will continue to rise to the challenges of producing rigorous and ethical empirical work in this area.

Appendix A

Supplemental information for
experimental evaluation of
observational methods

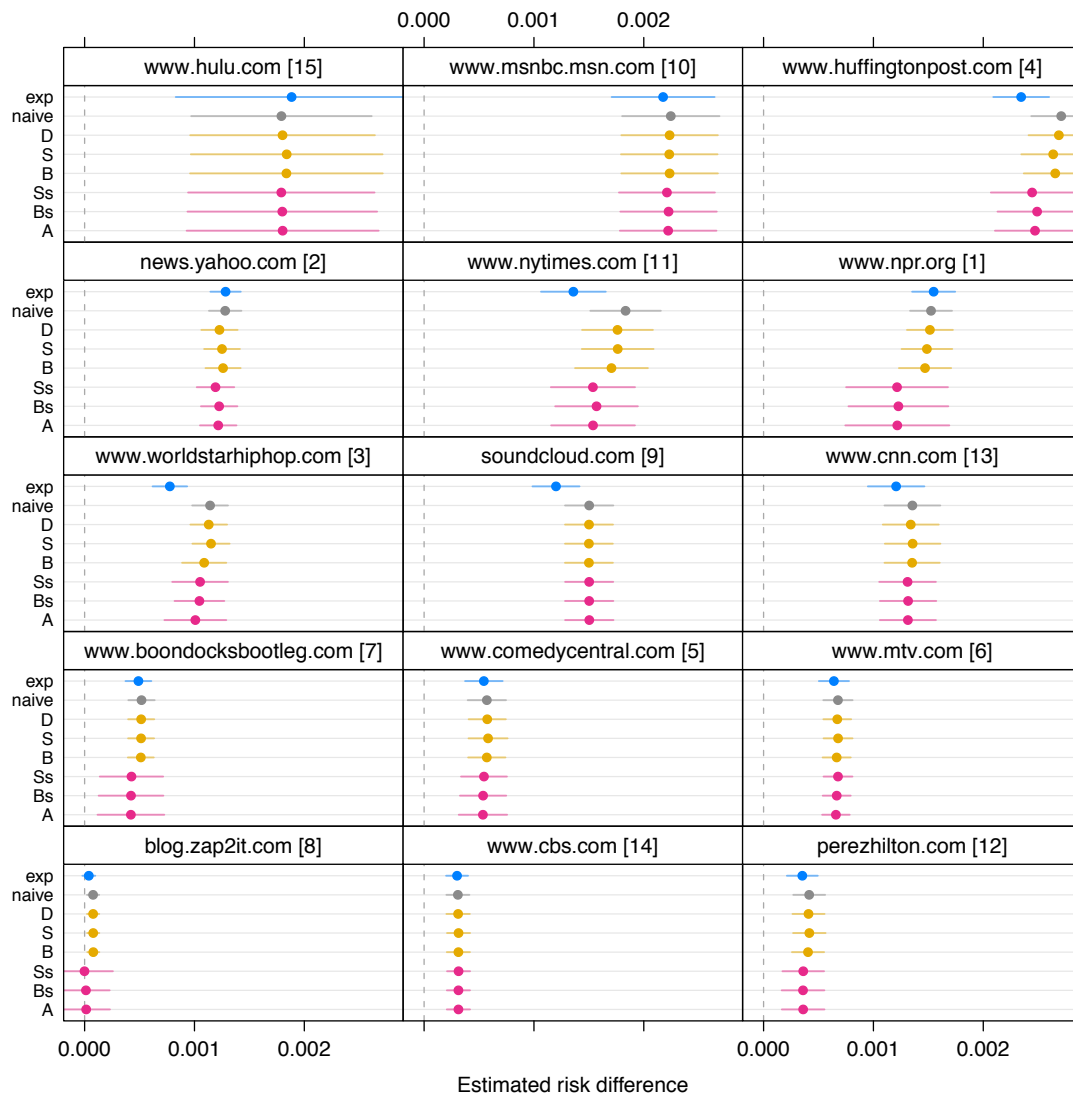


Figure A.1: Estimated peer effects via News Feed using stratification on propensity scores for the 15 domains with the largest number of exposed individual-URL pairs; numbers in brackets indicate this rank. Domains are sorted by the experimental estimate. Error bars are 95% bootstrap confidence intervals from reweighting both users and URLs.

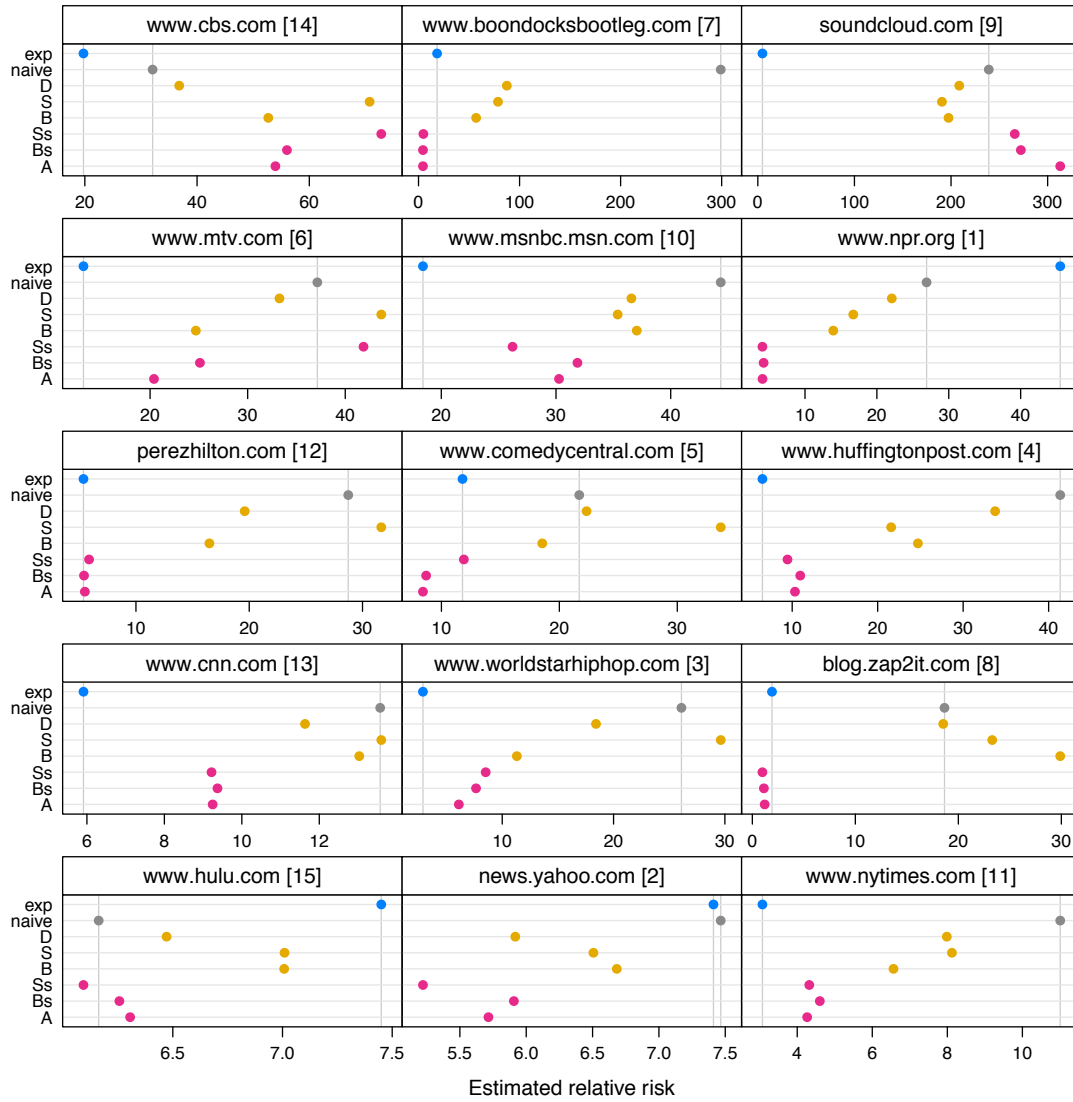


Figure A.2: Estimated peer effects in terms of relative risk for the 15 domains with the largest number of exposed individual-URL pairs; numbers in brackets indicate this rank. Note the very large experimental estimate for www.npr.org.

Appendix B

Supplemental information for
study of peer effects in cultural
rituals

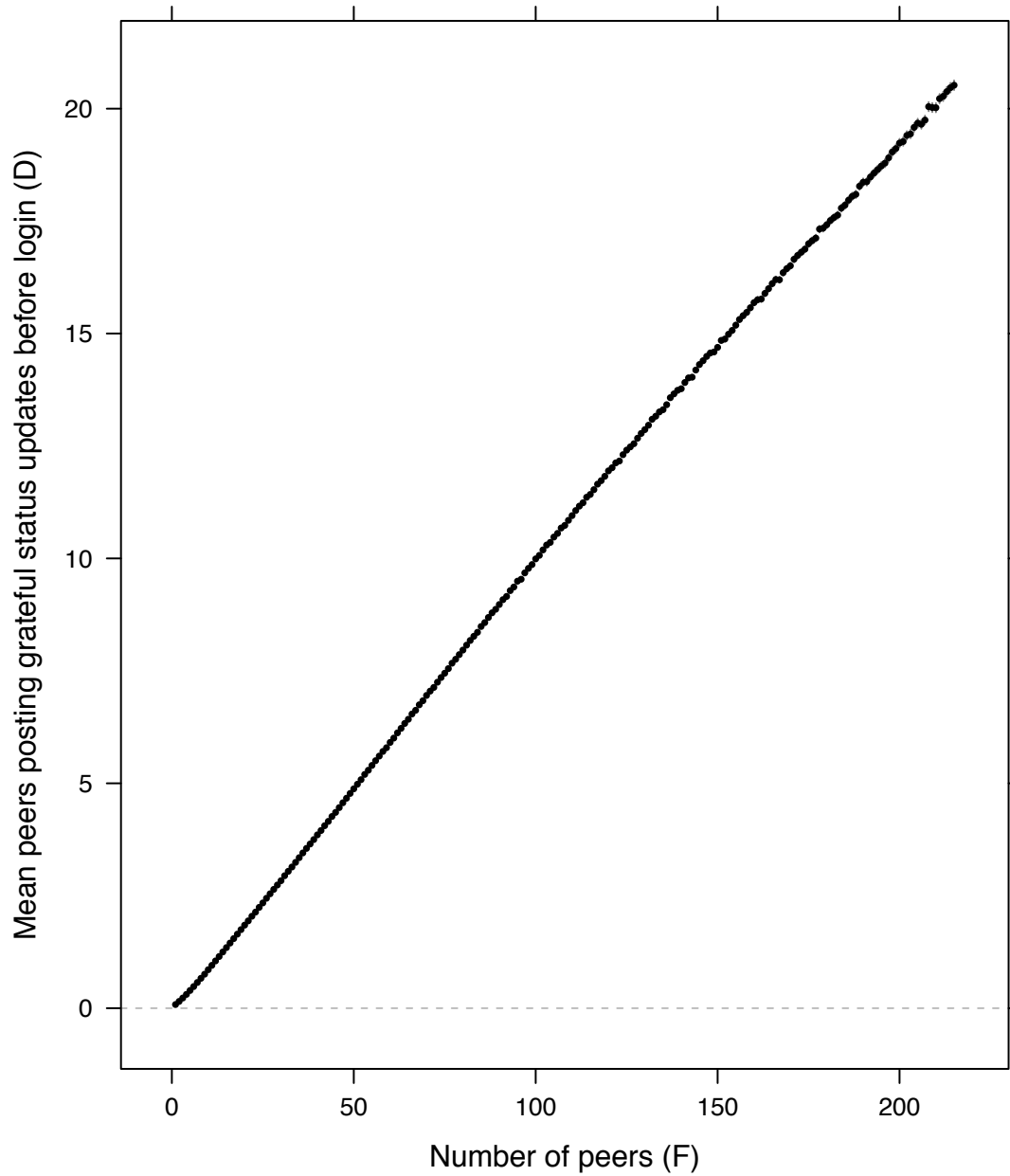


Figure B.1: Mean number of peers posting status updates before the ego logs in, D , as a function of observed number of peers, F . Perhaps surprisingly, the relationship is notably linear, given that it might be expected that peers of individuals for (a) log in at different times and (b) have different number of total peers might have systematically heterogeneous patterns of peer behavior. Error bars are 95% confidence intervals.

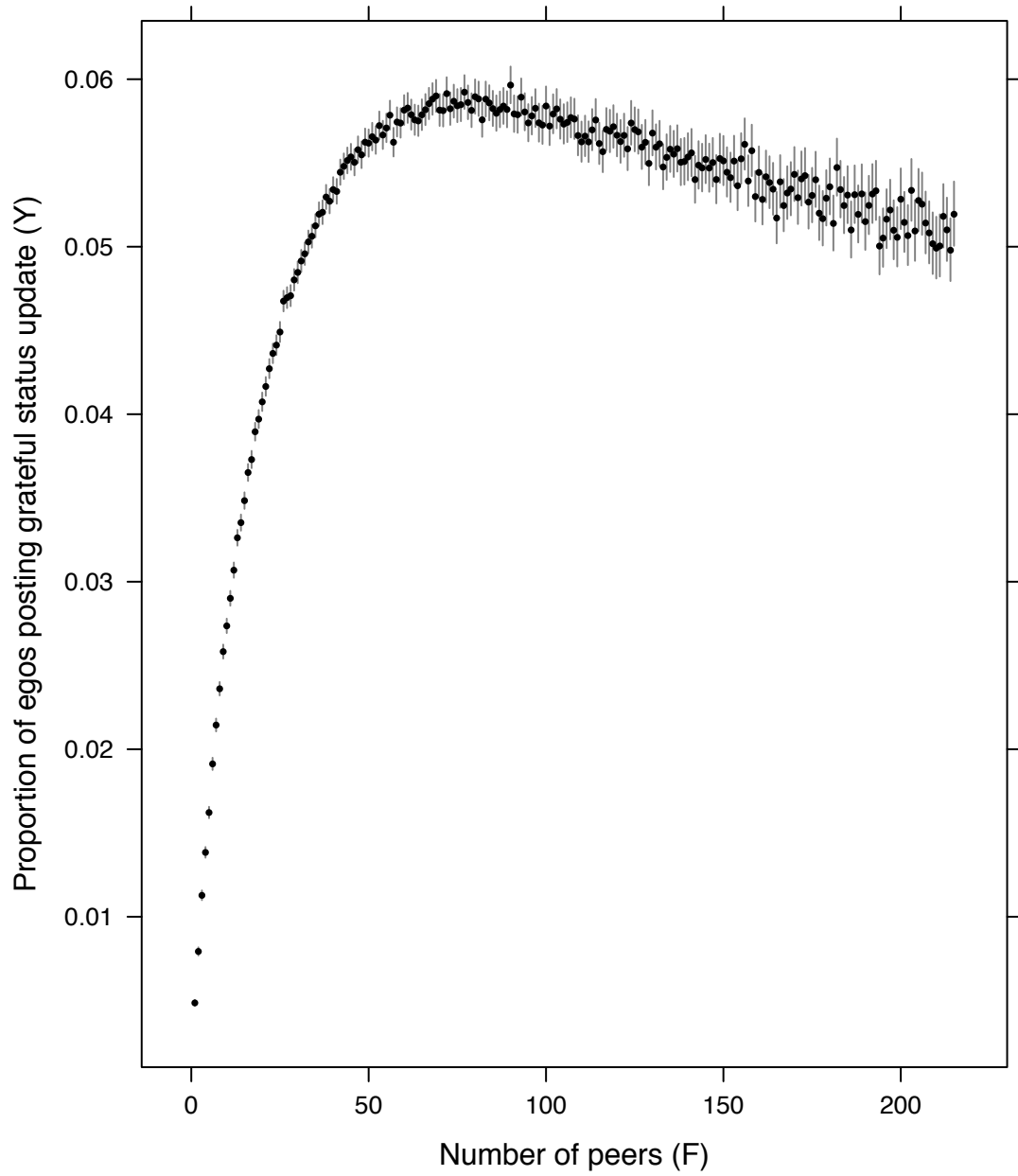


Figure B.2: Probability of ego behavior as a function of observed number of peers F . This association is notably nonlinear, suggesting that care should be taken that Z is made ignorable by appropriately conditioning on F . Error bars are 95% confidence intervals.

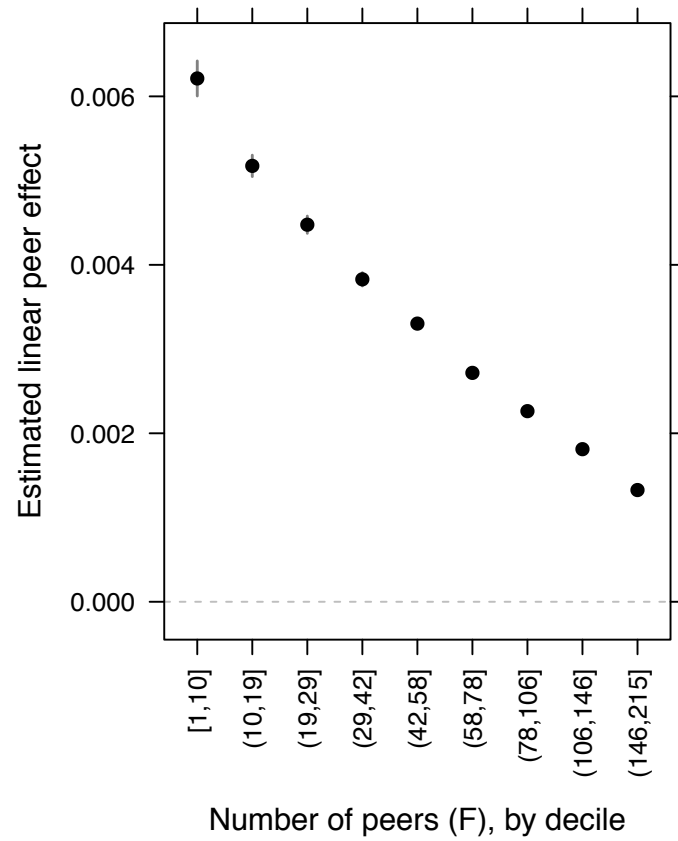


Figure B.3: Estimated peer effects from a purely observational analysis regressing Y on D and indicators for each value of F . The indicators for F are centered and interacted with D so as to, under ignorability, estimate the ATE rather than a precision-weighted average. Error bars are 95% confidence intervals.

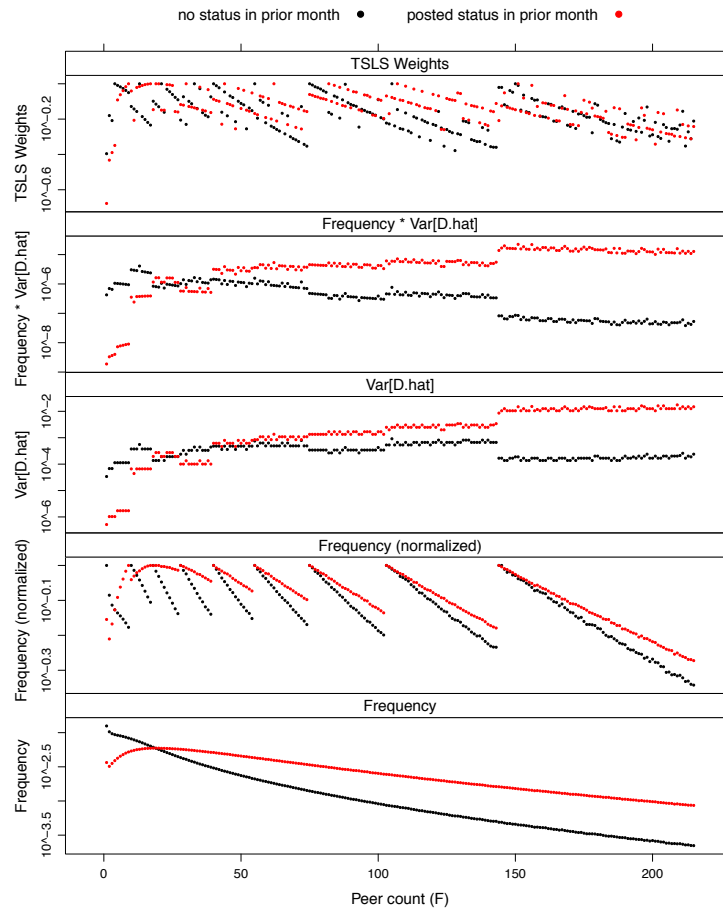


Figure B.4: Two-stage least squares estimates are weighted by the variance of the fitted values for D from the first stage. This figure displays factors contributing to the weighting function. The normalized frequency and weights panels are normalized within each decile of F to illustrate the relative weights for each estimate in the IV analysis. Comparison of those two panels illustrates that egos with larger values of F are up-weighted in the TSLS estimation.

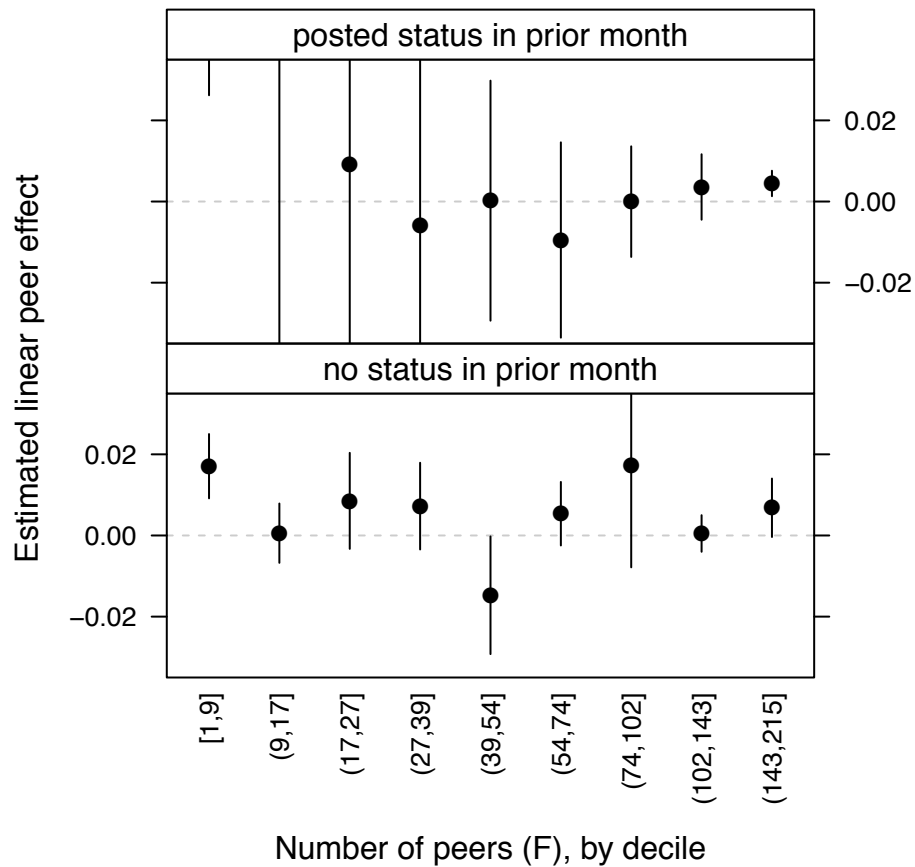


Figure B.5: Instrumental variable estimates of peer effects by deciles of number of peers, F , and prior posting of status updates within the first hour after login. These estimates are for an alternative model not described in the text. This model includes an interaction of Z with a piecewise linear function of F with knots at the even percentiles of F . This increases the number of instruments and thus weak instrument bias. The results are qualitatively similar to the model reported in the text. Error bars are 95% confidence intervals.

Bibliography

- Abadie, A. and Imbens, G. W. (2008). On the failure of the bootstrap for matching estimators. *Econometrica*, 76(6):1537–1557.
- Angelucci, M. and De Giorgi, G. (2009). Indirect effects of an aid program: How do cash transfers affect ineligibles consumption? *American Economic Review*, 99(1):486508.
- Angrist, J. and Imbens, G. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431442.
- Angrist, J. and Pischke, J. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *National Bureau of Economic Research Working Paper Series*, No. 15794. published as.
- Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous regressors. *Journal of Business & Economic Statistics*, 19(1):2–28.
- Angrist, J. D. and Imbens, G. W. (1999). Comment on James J. Heckman, “Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations”. *The Journal of Human Resources*, 34(4):823–827.

- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434).
- Aral, S. (2011). Commentary — Identifying social influence: A comment on opinion leadership and social contagion in new product diffusion. *Marketing Science*, 30(2):217–223.
- Aral, S., Muchnik, L., and Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549.
- Aronow, P. and Samii, C. (2011). Estimating average causal effects under general interference. Manuscript.
- Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, 70(9):1–70.
- Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., and Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *Journal of the Learning Sciences*, 17(1):102141.
- Baker, J. W. (2009). *Thanksgiving: The Biography of an American holiday*. University of New Hampshire Press.
- Baker, S. M. and Petty, R. E. (1994). Majority and minority influence: Source-position imbalance as a determinant of message scrutiny. *Journal of Personality and Social Psychology*, 67(1):5–19.

- Bakshy, E., Eckles, D., Yan, R., and Rosenn, I. (2012a). Social influence in social advertising: Evidence from field experiments. In *Proceedings of the ACM conference on Electronic Commerce*. ACM.
- Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. (2012b). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web, WWW '12*, pages 519–528. ACM.
- Bala, V. and Goyal, S. (1998). Learning from neighbours. *The Review of Economic Studies*, 65(3):595–621.
- Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176.
- Berg, N. and Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas*, 18(1):133166.
- Blume, L. E. (1995). The statistical mechanics of Best-Response strategy revision. *Games and Economic Behavior*, 11(2):111–145.
- Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, 36(1):75–111.
- Bramoulle, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55.
- Brennan, R. L., Harris, D. J., and Hanson, B. A. (1987). The bootstrap and other procedures for examining the variability of estimated variance components in testing contexts. Technical report, American College Testing Program.

- Burke, M., Kraut, R., and Joyce, E. (2010). Membership claims and requests: Conversation-Level newcomer socialization strategies in online groups. *Small Group Research*, 41(1):4–40.
- Burke, M., Marlow, C., and Lento, T. (2009). Feed me: Motivating newcomer contribution in social network sites. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 945–954, Boston, MA, USA. ACM.
- Cacioppo, J. T., Petty, R. E., Kao, C. F., and Rodriguez, R. (1986). Central and peripheral routes to persuasion: An individual difference perspective. *Journal of Personality and Social Psychology*, 51(5):1032–1043.
- Cai, J. (2011). Social networks and the decision to insure: Evidence from randomized experiments in china. Manuscript.
- Card, D., Mas, A., and Rothstein, J. (2008). Tipping and the dynamics of segregation. *The Quarterly Journal of Economics*, 123(1):177–218.
- Carrell, S. E., Fullerton, R. L., and West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3):439–464.
- Carrell, S. E., Hoekstra, M., and West, J. E. (2011a). Is poor fitness contagious?: Evidence from randomly assigned friends. *Journal of Public Economics*, 95(7–8):657–663.
- Carrell, S. E., Sacerdote, B. I., and West, J. E. (2011b). From natural variation to optimal policy? The lucas critique meets peer effects. *National Bureau of Economic Research Working Paper Series*, No. 16865.

- Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197.
- Centola, D. and Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113:702–734.
- Chaiken, S. and Chen, S. (1999). The heuristic–systematic model in its broader context. In Chaiken, S. and Trope, Y., editors, *Dual-process Theories in Social Psychology*, pages 73–96. Guilford Press.
- Chaiken, S., Giner-Sorolla, R., and Chen, S. (1996). Beyond accuracy: Defense and impression motives in heuristic and systematic information processing. In *The psychology of action: Linking cognition and motivation to behavior*, pages 553–578. New York, NY, US: Guilford Press.
- Chen, S., Duckworth, K., and Chaiken, S. (1999). Motivated heuristic and systematic processing. *Journal of Personality and Social Psychology*, 47:245–287.
- Christakis, N. A. and Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *N Engl J Med*, 357(4):370–379.
- Christakis, N. A. and Fowler, J. H. (2008). The collective dynamics of smoking in a large social network. *N Engl J Med*, 358(21):2249–2258.
- Cohen-Cole, E. and Fletcher, J. M. (2008a). Detecting implausible social network effects in acne, height, and headaches: Longitudinal analysis. *BMJ*, 337(2):2533–2533.
- Cohen-Cole, E. and Fletcher, J. M. (2008b). Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics*, 27(5):1382–1387.

- Coleman, J., Katz, E., and Menzel, H. (1957). The diffusion of an innovation among physicians. *Sociometry*, 20(4):253–270.
- Currarini, S., Jackson, M. O., and Pin, P. (2010). Identifying the roles of race-based choice and chance in high school friendship network formation. *Proceedings of the National Academy of Sciences*, 107(11):4857–4861.
- Deutsch, M. and Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, 51(3):629–636.
- Duflo, E. and Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *Quarterly Journal of Economics*, 118(3):815–842.
- Durlauf, S. N. and Ioannides, Y. M. (2010). Social interactions. *Annual Review of Economics*, 2(1):451–478.
- Easley, D. and Kleinberg, J. (2010). *Networks, Crowds, and Markets*. Cambridge University Press.
- Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, 25(5):603–637.
- Fletcher, J. M. (2010). Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Economics*, 19(4):466–484.
- Fodor, J. A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, 3(1):63–73.

- Friend, R., Rafferty, Y., and Bramel, D. (1990). A puzzling misinterpretation of the asch “conformity” study. *European Journal of Social Psychology; European Journal of Social Psychology*, 20(1):29–44.
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge University Press.
- Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3):472–482.
- Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443.
- Granovetter, M. and Soong, R. (1983). Threshold models of diffusion and collective behavior. *The Journal of Mathematical Sociology*, 9(3):165–179.
- Greiner, D. J. and Rubin, D. B. (2010). Causal effects of perceived immutable characteristics. *Review of Economics and Statistics*, 93(3):775–785.
- Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11(1):1–12.
- Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4):605–654.
- Heinsman, D. T. and Shadish, W. R. (1996). Assignment methods in experimentation: When do nonrandomized experiments approximate answers from randomized experiments? *Psychological Methods; Psychological Methods*, 1(2):154–169.

- Hill, J. (2008). Comment. *Journal of the American Statistical Association*, 103(484):1346–1350.
- Hill, J. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.
- Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models. *Sociological Methodology*, 18:449–484.
- Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1):4–29.
- Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press.
- Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.
- Kallgren, C. A., Reno, R. R., and Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and Social Psychology Bulletin*, 26(8):1002–1012.
- Kelman, H. C. (1958). Compliance, identification, and internalization: Three processes of attitude change. *The Journal of Conflict Resolution*, 2(1):51–60.
- Kelman, H. C. (1961). Processes of opinion change. *Public Opinion Quarterly*, 25(1):57–78.
- Kermack, W. and McKendrick, A. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London*, 115:700–721.

- Kossinets, G. and Watts, D. J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology*, 115(2):405–450.
- Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Pers Soc Psychol Bull*, 21(1):58–75.
- Kremer, M. and Levy, D. (2008). Peer effects and alcohol use among college students. *The Journal of Economic Perspectives*, 22(3):189–3A.
- LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, 76(4):604–620.
- Lasswell, H. (1948). The structure and function of communication in society. In Bryson, L., editor, *The Communication of Ideas*, volume 1, pages 117–130. University of Illinois Press, Urbana, Illinois.
- Latane, B. (1996). Dynamic social impact: The creation of culture by communication. *Journal of Communication*, 46(4):13–25.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., and Van Alstyne, M. (2009). Computational social science. *Science*, 323(5915):721–723.
- Leibenstein, H. (1950). Bandwagon, snob, and veblen effects in the theory of consumers’ demand. *The Quarterly Journal of Economics*, 64(2):183–207.
- Lunceford, J. K., Davidian, M., Lunceford, J. K., and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, 23(19):2937–2960.

- Lundborg, P. (2006). Having the wrong friends? Peer effects in adolescent substance use. *Journal of Health Economics*, 25(2):214–233.
- Manski, C. F. (1993). Identification problems in the social sciences. *Sociological Methodology*, 23:1–56.
- Manski, C. F. (2000). Economic analysis of social interactions. *The Journal of Economic Perspectives*, 14(3):115–136.
- Manski, C. F. (2008). *Identification for Prediction and Decision*. Harvard University Press.
- Martin, R. and Hewstone, M. (2002). Conformity and independence in groups: Majorities and minorities. In Hogg, M. A. and Tindale, R. S., editors, *Blackwell Handbook of Social Psychology: Group Process*, pages 209–234. Blackwell Publishing.
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444.
- Miguel, E. and Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.
- Moffitt, R. A. (2001). Policy interventions, low-level equilibria, and social interactions. In Durlauf, S. N. and Young, H. P., editors, *Social Dynamics*, page 4582. MIT Press.
- Montanari, A. and Saberi, A. (2010). The spread of innovations in social networks. *Proceedings of the National Academy of Sciences*, 107(47):20196 –20201.

- Moore, H. T. (1921). The comparative influence of majority and expert opinion. *The American Journal of Psychology*, 32(1):16–20.
- Morgan, S. L. and Harding, D. J. (2006). Matching estimators of causal effects prospects and pitfalls in theory and practice. *Sociological Methods & Research*, 35(1):3–60.
- Morgan, S. L. and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press.
- Moscovici, S., Lage, E., and Naffrechoux, M. (1969). Influence of a consistent minority on the responses of a majority in a color perception task. *Sociometry*, 32(4):365–380.
- Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, 6(1990):46247.
- Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*. MIT Press.
- Owen, A. B. (2007). The pigeonhole bootstrap. *The Annals of Applied Statistics*, 1(2):386–411.
- Owen, A. B. and Eckles, D. (2012). Bootstrapping data arrays of arbitrary order. *The Annals of Applied Statistics*. Forthcoming.
- Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.
- Pearl, J. (2009a). Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146.
- Pearl, J. (2009b). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

- Petty, R. E. and Briol, P. (2008). Persuasion: From single to multiple to metacognitive processes. *Perspectives on Psychological Science*, 3(2):137–147.
- Petty, R. E. and Wegener, D. T. (1999). The elaboration likelihood model: Current status and controversies. In Chaiken, S. and Trope, Y., editors, *Dual-process theories in social psychology*, pages 41–72. Guilford Press, New York.
- Pool, G. J. and Schwegler, A. F. (2007). Differentiating among motives for norm conformity. *Basic and Applied Social Psychology*, 29(1):47–60.
- Rogers, E. M. (2003). *Diffusion of Innovations*. Simon and Schuster, 5th edition.
- Rosenbaum, P. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society. Series A (General)*, 147(5):656–666.
- Rosenbaum, P. R. and Rubin, D. B. (1984a). On the nature and discovery of structure: Comment. *Journal of the American Statistical Association*, 79(385):26–28.
- Rosenbaum, P. R. and Rubin, D. B. (1984b). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387):516–524.
- Rosenzweig, M. and Wolpin, K. I. (2000). Natural natural experiments. *Journal of Economic Literature*, 38(4):827–874.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

- Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, 127(8 Part 2):757–763.
- Rubinstein, A. (2006). *Lecture Notes in Microeconomic Theory: The Economic Agent*. Princeton University Press.
- Ryan, B. and Gross, N. C. (1943). The diffusion of hybrid seed corn in two iowa communities. *Rural Sociology*, 8(1):1524.
- Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, 116(2):681–704.
- Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856.
- Schafer, J. L. and Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, 13(4):279–313.
- Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology*, 1(2):143–186.
- Shadish, W. R., Clark, H. H., and Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, 103(484):1334–1344.
- Shalizi, C. R. and Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40(2):211–239.

- Sherif, M. (1936). *The Psychology of Social Norms*. Harper, New York.
- Shriver, S. K., Nair, H., and Hofstetter, R. (2011). Social ties and user-generated content: Evidence from an online social network. Technical Report 2083, Stanford Graduate School of Business.
- Smith, E. R. and DeCoster, J. (2000). Dual-Process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4(2):108–131.
- Thomas, A. C. and Blitzstein, J. K. (2011a). Valued ties tell fewer lies, II: Why not to dichotomize network edges with bounded outdegrees. Manuscript. <http://arxiv.org/abs/1101.2228>.
- Thomas, A. C. and Blitzstein, J. K. (2011b). Valued ties tell fewer lies: Why not to dichotomize network edges with thresholds. Manuscript. <http://arxiv.org/abs/1101.0788>.
- Toelch, U., Bruce, M. J., Meeus, M. T., and Reader, S. M. (2010). Humans copy rapidly increasing choices in a multiarmed bandit problem. *Evolution and Human Behavior*, 31(5):326–333.
- Tucker, C. (2008). Identifying formal and informal influence in technology adoption with network externalities. *Management Science*, 54(12):2024–2038.
- Turner, J., Hogg, M., Oakes, P., Reicher, S., and Wetherell, M. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. Blackwell, Oxford.
- Valente, T. W. (1995). *Network models of the diffusion of innovations*. Hampton Press.

- Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. *Social Networks*, 18(1):69–89.
- VanderWeele, T. J. (2011). Sensitivity analysis for contagion effects in social networks. *Sociological Methods & Research*, 40(2):240–255.
- Wallendorf, M. and Arnould, E. J. (1991). “We gather together”: Consumption rituals of Thanksgiving Day. *Journal of Consumer Research*, 18(1):13–31.
- Walther, J., Van Der Heide, B., Kim, S., Westerman, D., and Tong, S. (2008). The role of friends appearance and behavior on evaluations of individuals on facebook: Are we known by the company we keep? *Human Communication Research*, 34(1):28–49.
- Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771.
- Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology*, 51:539–570.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20(7):557–585.
- Young, H. P. (2009). Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review*, 99(5):1899–1924.
- Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics*, 85(1):9–23.