Estimating peer effects in networks with peer encouragement designs

Dean Eckles[a,b,1], René F. Kizilcec[b,c], and Eytan Bakshy[b]

[a]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Facebook, Menlo Park, CA 94025; and [c]Department of Communication, Stanford University, Stanford, CA 94305

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved April 15, 2016 (received for review June 15, 2015)

Peer effects, in which the behavior of an individual is affected by the behavior of their peers, are central to social science. Because peer effects are often confounded with homophily and common external causes, recent work has used randomized experiments to estimate effects of specific peer behaviors. These experiments have often relied on the experimenter being able to randomly modulate mechanisms by which peer behavior is transmitted to a focal individual. We describe experimental designs that instead randomly assign individuals' peers to encouragements to behaviors that directly affect those individuals. We illustrate this method with a large peer encouragement design on Facebook for estimating the effects of receiving feedback from peers on posts shared by focal individuals. We find evidence for substantial effects of receiving marginal feedback on multiple behaviors, including giving feedback to others and continued posting. These findings provide experimental evidence for the role of behaviors directed at specific individuals in the adoption and continued use of communication technologies. In comparison, observational estimates differ substantially, both underestimating and overestimating effects, suggesting that researchers and policy makers should be cautious in relying on them.

social interactions | social networks | causal inference | experimental design

Social interactions among people enable the spread of information, preferences, and behavior, including technology adoption. Despite the unprecedented availability of detailed information on human interactions, credible identification of how individuals affect each other has been difficult. Many of the empirical studies that estimate these peer effects rely on analyzing observational (i.e., nonexperimental) data (e.g., refs. 1 and 2). These methods can incorrectly "detect" peer effects in their absence (3–5) and can substantially overestimate them (6). There are many causes of correlated behaviors in networks that make it difficult to identify peer effects, including selective tie formation [i.e., homophily (7)], unobserved correlated external causes, and prior peer effects (4, 8, 9). Faced with these challenges, observational studies of peer effects are sometimes described as tentatively providing evidence of peer effects (cf. refs. 10 and 11) or as providing upper bounds, rather than point estimates, for peer effects (6).

This paper presents designs for randomized experiments for estimating peer effects in social networks that overcome common challenges to credible identification. We conducted a large field experiment on Facebook that implements a peer encouragement design to estimate peer effects in the use of communication technologies. In particular, many people share information, personal media, or other content via online social networks. Most of these services allow them to receive feedback from their peers in the form of comments on their post and expressions of approval (or disapproval). How does receiving more or less of this feedback from peers affect use of these technologies? Decision makers benefit from knowing the value of receiving social feedback, relative to other potential actions, as this informs the design of interfaces for giving feedback. For social scientists, precisely estimating the effects of feedback is important for, e.g., understanding network effects in the adoption and continued use of communication technologies.

There is some theoretical and empirical support for expecting substantial peer effects in initial adoption and use of communication technologies. Individuals' utilities from using such technologies usually depend on peer adoption decisions, as this determines who can be communicated with and the consequences of communication. Prior work on Facebook specifically (12, 13), and other related technologies (14, 15) has provided observational and quasi-experimental evidence for peer effects in initial adoption, content production, and sustained use. Other prior observational research has found that receiving feedback (e.g., comments, "likes") is associated with higher rates of sharing; in particular, new users who receive comments on their photos tend to share more photos in the future (10). However, in the presence of confounding due to homophily and common external causes, these prior observational results are expected to overstate (or otherwise misstate) the relationship between receiving feedback and subsequent behavior if interpreted causally.

Peer Encouragement Designs

Randomized experiments are one appealing way to identify peer effects in the presence of unknown confounding (16–18). Although directly randomizing the behavior of existing peers in realistic settings is generally not possible or desirable, multiple experimental designs for learning about peer effects appear in the literature. Social psychologists have used inauthentic, confederate peers since the 1950s (19, 20), often in artificial (e.g., laboratory) settings. Other studies have induced random variation in the process of tie or group formation (21–25). Although these approaches have been successful at answering some important questions in the social sciences, it is often not possible for such designs to credibly answer questions about either existing peers or effects of specific peer behaviors.

The widespread adoption of online social networks has facilitated in situ studies of the effects of peers' behaviors on individual behavior. Much of the experimental work in this area has used mechanism designs, which directly modulate mechanisms (or channels) by which information about peer behavior is optionally or nondeterministically

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Drawing Causal Inference from Big Data," held March 26–27, 2015, at the National Academies of Sciences in Washington, DC. The complete program and video recordings of most presentations are available on the NAS website at www.nasonline.org/Big-data.

Author contributions: D.E., R.F.K., and E.B. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

Conflict of interest statement: The authors were all employed by Facebook while conducting this research. D.E. and E.B. have significant financial interests in Facebook.

This article is a PNAS Direct Submission.

Data deposition: Analysis code and aggregate statistics for reproducing the main results (Figs. 3–5) are archived at the Harvard Dataverse Network (dx.doi.org/10.7910/DVN/ELUQVD).

[1]To whom correspondence should be addressed. Email: dean@deaneckles.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1511201113/-/DCSupplemental.

transmitted to a focal individual (ego) through the network (18, 26–28). For example, Aral and Walker (26) randomize which peers are sent viral messages to adopt a product, and Bakshy et al. (18) randomize the number of personalized social cues in advertisements. The causal directed acyclic graph (DAG) (29) shown in Fig. 1*A* illustrates a mechanism design with binary peer behaviors and ego behavior. When these designs involve enabling/disabling a mechanism of peer effects, they allow estimating an average treatment effect on the treated (ATT)—the effect of exposure to a peer behavior for those who would normally be exposed; if the mechanism is normally deterministic, this is also an ATT for the peer behavior, not just for exposure. Despite their advantages, mechanism designs are often not possible or practical in many empirical settings, such as when the mechanism is deterministic (i.e., information about a peer's behavior is always transmitted to the ego, such as feedback in an online social network).

Encouragement Designs. We develop and illustrate a variation on randomized encouragement designs for identifying effects in networks. Encouragement designs (30) are widely used by social and biomedical scientists when interested in the effects of behaviors not directly controlled by the experimenter. Units are randomly assigned to an encouragement Z_i, and the endogenous behavior of interest D_i and the outcome Y_i are measured. For example, in educational contexts, one may encourage children to watch a particular educational program (31) or prepare for tests (32). Not all parents or children may follow through with such interventions, but it is still possible to analyze the causal effect of the intervention for those who are induced to use the educational materials by the randomized encouragement (i.e., for compliers). For this purpose, the encouragement is treated as an instrumental variable (IV); that is, it is assumed that the encouragement only affects outcomes by affecting the intermediate, endogenous behavior of interest. This complete mediation or exclusion restriction can be stated as follows. Define the potential outcomes for Y_i and D_i as functions of the encouragement and the behavior, $Y_i : \mathbb{D} \times \mathbb{Z} \to \mathbb{Y}$ and $D_i : \mathbb{Z} \to \mathbb{D}$.

Assumption 1. (Exclusion restriction). *Suppose* $Y_i(d_i, z_i) = Y_i(d_i, z_i')$ *for all* $d_i \in \mathbb{D}$ *and* $z_i, z_i' \in \mathbb{Z}$, *so we can uniquely define* $Y_i(d_i)$.

[Combining this assumption with the random assignment of Z_i, some authors (e.g., ref. 33), write $Y_i(d_i), D_i(z_i) \perp\!\!\!\perp Z_i$ for all $d_i \in \mathbb{D}$ and $z_i, Z_i \in \mathbb{Z}$. The exclusion restriction is then combined with either parametric assumptions about $Y_i(\cdot)$ or nonparametric assumptions about $D_i(\cdot)$ to identify effects of D_i on Y_i (34, 35); both are discussed in *Model*. Here and elsewhere, we use capital letters for random variables and lowercase letters for fixed values. We retain subscripts for units even in the latter case, as those without subscripts denote n vectors.

Encouragement Designs in Groups and Networks. Peer encouragement designs randomize an individual's peers to conditions that increase or decrease the probability of those peers performing a specific behavior. One may then examine how this shock to peer behaviors "spills over" to the behaviors of focal individuals. Furthermore, these designs can provide point estimates of the effect of peer behavior on ego behavior (i.e., peer effects) by using encouragements to a specific behavior and assuming that the only effect of peer assignment to these encouragements on ego behavior is via that specific peer behavior. The causal DAG in Fig. 1*B* illustrates a peer encouragement design with binary encouragements, peer behaviors, and ego behavior. In this example, the encouragement causes one peer to adopt the specific behavior, which in turn causes the ego to adopt. Given the assumptions encoded in this DAG, peer encouragement is an IV, and we can estimate the effect of the behavior of peers, as caused by the encouragement to adopt, on ego behavior.

Plausibility of the Exclusion Restriction. This DAG encodes an exclusion restriction (Assumption 1): All effects of the peer encouragement on ego behavior occur via changes to peer behavior. In standard encouragement designs, the randomized encouragement Z_i, endogenous behavior D_i, and outcome Y_i are all defined as direct interventions on or measurements of the same individual, often making this assumption implausible because the encouragement may affect that individual in many ways (30, 36, 37). For example, parents encouraged to watch Sesame Street with their children (31) may modify their child-rearing in many ways. In peer encouragement designs, however, the exclusion restriction assumption is often particularly plausible for a structural reason: For a given ego whose outcome is observed, the encouragement is applied to other units—their peers. The ego usually does not directly observe or experience the encouragement; instead, it only affects the ego through peer behavior and, the researcher hopes to ensure, primarily through a small number of measured peer behaviors. We make the following two design recommendations—implemented in our empirical example—that can increase the plausibility of the exclusion restriction and increase statistical power.

First, provided the sample is sufficiently large, selecting a peer encouragement that is minimal may reduce the potential for reactance; we illustrate this point by comparison. Many designs that randomly assign treatment and estimate "spillovers" (i.e., interference or exogenous peer effects) (38, 39) can be understood and analyzed as peer encouragement designs. Recent work by economists and political scientists has examined such spillover effects within groups (40–44) or in a social network (45–47). In

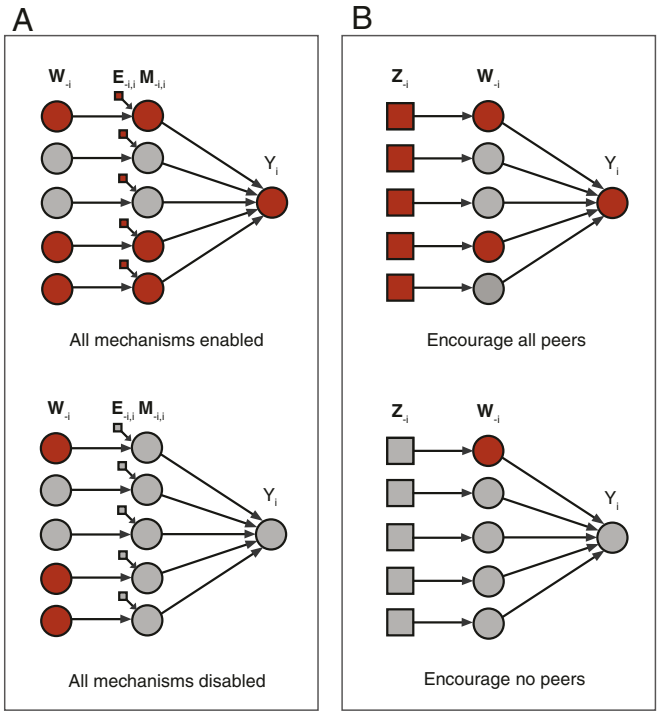

Fig. 1. Mechanism designs and peer encouragement designs for estimating peer effects, illustrated with binary variables. W_{-i} indicate peers' behaviors, and Y_i represent the behavior of a focal individual (ego). Variables are colored to represent example values under different random assignments (red = 1, gray = 0). (*A*) Mechanism designs modulate a channel by which peer effects occur, for example, by randomly enabling or disabling ($E_{-i,i}$) a particular mechanism ($M_{-i,i}$) by which a focal individual (i) is exposed to peer behavior ($-i$). (*B*) Peer encouragement designs use randomized encouragements to peers (Z_{-i}). All variables represented by circles may have other common causes not shown. Variables represented by squares are root nodes and are determined by random assignment.

some cases, researchers have attributed the estimated spillovers from treatment assignment to a specific peer behavior: In one study (41), employees were randomly assigned to encouragements to attend a retirement benefits fair. Among other analyses, Duflo and Saez (41) attribute spillover effects on retirement plan enrollment to effects of peer attendance at the fair. However, as the authors note, this encouragement may have directly affected always-attenders [e.g., via self-perception or crowd-out effects (48)], never-attenders (e.g., via salience of benefits), and their peers (e.g., via increased discussion of benefits). If, instead, the encouragement was unlikely to be remembered or even consciously perceived as an inducement, perhaps such violations of the exclusion restriction would be less likely to occur. Thus, peer encouragement designs could provide more credible peer effect estimates if the encouragement is a minimal "nudge" that may not warrant much conscious consideration.

A second design recommendation is, when appropriate for the research question, to use encouragements that are specific to particular directed edges, rather than encouraging a general, undirected behavior in peers. The experiments mentioned above generally use the number or fraction of assigned peers as the instrument. This instrument is then necessarily correlated for all egos in the same group or, more generally, who share peers. On the other hand, it is sometimes possible to encourage directed behaviors on particular edges; that is, an encouragement that induces a behavior from an alter j to an ego i. Such an encouragement could be randomly assigned at the level of the directed edge, or at the level of the target (i.e., the ego). In the latter ego-specific design, an ego i is randomly assigned to a peer encouragement condition Z_i, according to which all edges from any alter j to ego i are treated. That is, egos are randomly assigned to conditions that encourage their peers to engage in directed behaviors toward them; those same peers might be assigned to a different condition with respect to their other peers. In this ego-specific design, the instrument is no longer correlated within groups or in the network. This design choice can substantially change power; simulations on small-world networks demonstrate the ego-specific design reducing true SEs by 20% to over 90% (*SI Appendix*, *Simulations with Ego-Specific and General Designs*). Here we report on a large experiment in which the peer encouragement is a minimal change that causes a specific behavior directed at a particular ego.

Empirical Context and Data

Our empirical study examines the effects of receiving feedback from peers on Facebook. In particular, we examine feedback on socially shared content (posts) such as text, photos, videos, and links shared by egos. This content appears in the News Feeds of peers (friends), who may, in turn, provide feedback on these posts by providing comments on the post or clicking on the "Like" button. Individuals who receive feedback on a post may receive notifications immediately on Facebook, or via mobile notifications or email.

The design of a feedback interface poses a complex tradeoff: An interface that causes an ego's post to occupy more space in their peers' News Feeds may increase the likelihood that peers will provide feedback on the post; at the same time, such an interface may cut into peers' limited time and attention to view and interact with others' posts. To choose among interfaces, Facebook product teams frequently randomly assign some users to receive an alternative (often new) version of an interface to evaluate these alternatives; the data presented here arise from one such trial. In particular, we implemented a peer encouragement design that enables estimating the effects of receiving feedback when sharing content in social media.

Peers' responses to an ego's content (i.e., liking and commenting) are expected to vary with the user interface associated with that content when seen by peers. Egos were randomly assigned to conditions that encouraged their peers to provide feedback under different circumstances. There were two experimental factors that independently governed the display of egos' posts in peers' News Feeds (Fig. 2). First, the "encourage initiation" factor was relevant for posts without any feedback, and it determined whether the viewer would need to click "Comment" to display the textbox in which to write a comment or whether this textbox would be already visible. Second, the "conversation salience" factor was relevant for posts that had already received feedback, and it determined whether this existing feedback would be summarized numerically and displayed after a click or would already be visible (up to three comments shown by default). Thus, the encourage initiation factor should primarily cause the first feedback to occur at all or earlier, whereas the conversation salience factor should cause additional feedback. There were six possible conditions egos could be assigned to, resulting in a three (encourage initiation: always, sometimes, never) by two (conversation salience: high, low) design. (For the encourage initiation factor, the level sometimes was the default interface at the time: Posts displayed in the first position in News Feed would have the textbox shown, but posts appearing in other positions would not.)

This experiment thus is a peer encouragement design in which directed edges are treated according to an ego-specific random assignment: A particular person viewing their News Feed could see posts from multiple egos, which would be displayed according to the conditions to which each of those egos were assigned. We use this experiment to examine the effects of receiving feedback on how many posts egos make and how much feedback they give on others' posts. To establish a baseline for comparing effect sizes, we also estimate effects on how much they respond to feedback on their own posts. Feedback received is measured as the mean daily number of comments and likes received during the experimental period. All analysis is of deidentified data primarily consisting of counts of behaviors.

Model

For the main analysis, we work with log-transformations of the count variables (see *SI Appendix*, *Transformed and Untransformed Count Variables*). Let D_i be the logarithm of feedback received (likes and comments) by i during the experimental period and Y_i be the logarithm of one of the ego behaviors of interest. We aim to estimate effects of D_i on Y_i by using the random variation in D_i caused by assignment to the peer encouragement, Z_i. That is, we aim to summarize contrasts between potential

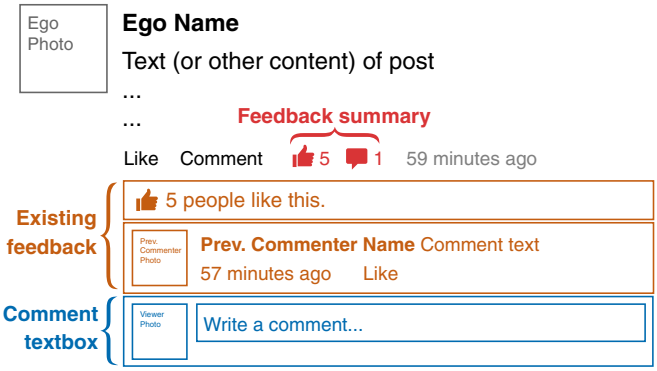

Fig. 2. Illustration of the feedback interfaces that would be used to display an ego's post to their peers according to which condition the ego was randomly assigned. For posts with feedback, the conversation salience factor determines whether the feedback is summarized numerically (red) or, instead, the existing feedback is shown (orange) with a textbox for writing a new comment (blue). In the low-salience case, a click on "Comment" or the feedback summary would display the existing feedback and comment textbox. For posts without feedback, the encourage initiation factor determines whether the comment textbox is shown by default (blue) or whether a click on "Comment" is needed to display it.

outcomes for different levels of feedback received, for example, some summary of $Y_i(d_i) - Y_i(d_i')$. Writing i's potential outcomes as functions only of the ith elements of an n-vector of feedback received requires two assumptions that specify the potential outcomes are constant in some inputs (i.e., specify level sets). First, this requires the exclusion restriction for IVs (Assumption 1). The minimal nature of the encouragement makes it plausible that it only affects egos by causing them to receive additional feedback; however, there may be effects of the feedback not captured by its quantity (e.g., content of comments, timing).

Additionally, already in writing $Y_i(d_i, z_i)$, we assume that the behaviors and assignments of all other units can be safely ignored—a "no interference" (49) or "individualistic treatment response" (50) assumption.

Assumption 2. (No interference). *Suppose that* $Y_i(d_i, d_{-i}, z_i, \mathbf{z}_{-i}) = Y_i(d_i, d'_{-i}, z_i, \mathbf{z}'_{-i})$ *for all* $d, d' \in \mathbb{D}^n$, $\mathbf{z}, \mathbf{z}' \in \mathbb{Z}^n$ *so that we can uniquely define* $Y_i(d_i, z_i)$.

This assumption is expected to be violated in this setting, even in our finite population. First, the units are interacting and make up a substantial portion of a single network. Second, the peer encouragement conditions would have different effects under a different global policy such that, e.g., peers were seeing all posts displayed according to the same interface rule. However, methods for statistical and causal inference in the presence of interference remain somewhat underdeveloped, especially for interference in a single network rather than within many isolated groups. We therefore work with the assumption that relevant nuisance interference is small compared with the effects of interest. For example, consider the assumption that this nuisance interference is no larger than than the effect of an increase c to feedback received.

Assumption 3. (Direct-effect-bounded interference). *Suppose that*

$$|Y_i(d_i, d_{-i}, z_i, \mathbf{z}_{-i}) - Y_i(d_i, d'_{-i}, z_i, \mathbf{z}'_{-i})| \le |Y_i(d''_i + c, d''_{-i}, z''_i, \mathbf{z}''_{-i}) - Y_i(d''_i, d''_{-i}, z''_i, \mathbf{z}''_{-i})|$$

for all $d, d', d'' \in \mathbb{D}^n$, $\mathbf{z}, \mathbf{z}', \mathbf{z}'' \in \mathbb{Z}^n$.

If, as in our main analysis, feedback received is modeled on a log scale, then, for $c = 1$, this assumes that any interference is smaller than the effect of multiplying feedback received by e (i.e., increasing feedback received by 172%); thus, sensitivity analysis based on such an assumption allows for very substantial interference. In *SI Appendix*, we combine this assumption with a specific model of local interference (50, 51) to conduct analyses quantifying the sensitivity of our results to nuisance interference. For simplicity, we now proceed with a model without nuisance interference.

In addition to Assumptions 1 and 2, there are multiple sets of assumptions that allow identification and estimation using peer encouragement conditions as IVs. One such assumption is that the effects of feedback received are (log–log) linear and constant; that is,

$$Y_i(d_i) - Y_i(0) = \gamma d_i.$$

In this case, two-stage least squares (TSLS) with multiple instruments simply increases precision in estimating γ; because both Y_i and D_i are on a logarithmic scale, γ is approximately the effect of a 1% increase in D_i in terms of percent change in Y_i. To estimate γ, we estimate the following two regression equations using TSLS:

$$Y = \mathbf{X}\mu + \gamma D + \varepsilon_i$$

$$D = \mathbf{X}\alpha + \mathbf{Z}\beta + \eta_i$$

where $\mathbf{X} = [\mathbf{S}\ \mathbf{C}]$ is a sparse $n \times 80{,}065$ matrix of (i) binary indicators for 64 strata formed by the quartiles of preexperiment feedback received, number of peers active on the web interface to Facebook, and preexperiment posting and (ii) binary indicators for 80,001 network clusters formed by graph partitioning, and $\mathbf{Z}$ is an $n \times k$ matrix of instruments, which are each binary indicators derived from the peer encouragement factors.

We expect the effects of feedback to be somewhat heterogeneous. "Marginal feedback," feedback that occurs (or does not occur) because of small changes, may be different from other feedback. Additionally, there may be heterogeneous effects of marginal feedback. For these reasons, we could adopt a nonparametric assumption on $D_i(\cdot)$ rather than a parametric assumption on $Y_i(\cdot)$: Each encouragement does not reduce feedback received for any egos.

Assumption 4. (Monotonicity). *Define* $h(\cdot) : \mathbb{Z} \to \{1, \ldots, k\}$ *to order the* k *values in* $\mathbb{Z}$ *such that* $j < l$ *implies* $\mathrm{E}[D_i | h(Z_i) = j] < \mathrm{E}[D_i | h(Z_i) = l]$. *With probability 1,* $D_i(z_i) - D_i(z_i') \ge 0$ *for all* $i \in \mathbb{P}_{egos}$, *where* $h(z_i) > h(z_i')$.

Then TSLS using binary indicators formed from the levels of $h(Z_i)$ estimates a weighted average of estimators using a single binary indicator (ref. 33, theorem 2), each of which estimates an average causal response (ACR), which is a weighted average of effects of changes in increments of D_i. Because $D_i = g(D_i^*)$ is transformed from its original, skewed count distribution, this means that the weights for a change to $g(d_i^*)$ from $g(d_i^* - 1)$ in this average are the normalized product of a difference in cumulative distribution functions for D_i at $g(d_i^* - 1)$ for that instrument and $g(d_i^*) - g(d_i^* - 1)$; see *SI Appendix*, *Transformed and Untransformed Count Variables*. Our main results use a first stage without interactions between the two factors, so this simple theorem does not directly apply to that model. However, Lochner and Moretti (ref. 52, proposition 2) show that TSLS nonetheless estimates a weighted average of the single instrument estimands. This weighting function is shown in *SI Appendix*, Fig. S7. We test the choice of this first-stage model and show in Fig. 4 that the results are not affected by instead using data-driven shrinkage and selection with the lasso.

Results

We first examine the effects of the peer encouragements on feedback received (i.e., first-stage effects). Both encouragement factors cause peers to comment on and like posts by egos, such that these factors increase (geometric) mean feedback received by 0.2–1.3% (Fig. 3*A*), $F(3, 4.9e7) = 519$, $p <$ 1e-12. Adding the two interaction terms for these factors did not significantly improve fit, $F(2, 4.9e7) = 0.23$, $p = 0.80$. As expected, the encourage initiation factor shifts the lower end of the distribution of feedback received more, compared with the conversation salience factor (Fig. 3*B*).

This randomly induced variation in feedback received allows us to estimate effects of receiving feedback on multiple ego behaviors. We focus on results from a first-stage specification as in Fig. 3*A*, with all three main effects (black points in Fig. 4).

Receiving additional feedback is expected to have the largest effects on "reply" behaviors by the ego, such as commenting on their own posts and liking comments on their posts. We estimate large effects of receiving feedback on both of these ego behaviors, such that a 10% increase in feedback received causes a 9.6% increase in comments (self) and a 10.5% increase in likes (self). Although unsurprising, these estimates can help put the magnitude of effects on other ego behaviors in perspective.

Effects on other ego behaviors are more important for understanding the spread of feedback and sharing behaviors. Receiving additional feedback also causes egos to give others more feedback, in terms of both likes and comments separately: Receiving 10% more feedback causes egos to give others 1.1% more likes and 1.1% more comments. Thus, causing one individual to receive more feedback will cause them to give more feedback to their peers, potentially creating desirable feedback loops. As expected, these effects are substantially smaller than

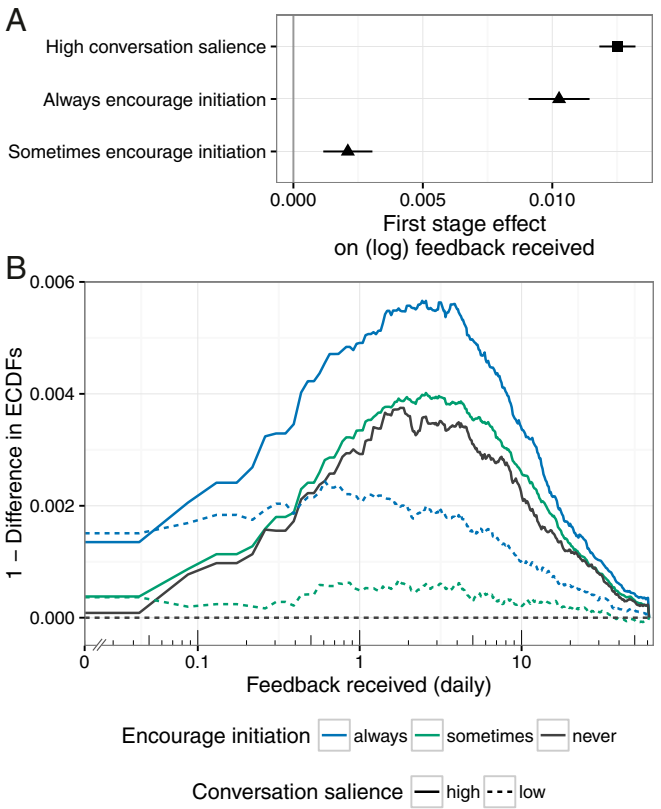

Fig. 3. Effects of the encouragements on feedback received (first stage). (*A*) First-stage average effects. Points are coefficient estimates for the effects of the conversation salience (circle) and encourage initiation (triangles) factors on (log) feedback received, where the base condition is low conversation salience and never encourage initiation. Error bars are 95% network adjacency- and cluster-robust confidence intervals. (*B*) Effects on the distribution of feedback received, computed as a difference in the empirical cumulative distribution functions (ECDFs) of feedback received. Again, with the lowest-feedback condition (never/low) as the baseline, each line represents the difference in probability that daily feedback received is at least the value on the *x* axis. The encourage initiation factor, which has its immediate effects only when a post has no feedback, has larger effects at the low end of the feedback distribution, whereas the conversation salience factor produces shifts at the high end. These differences use poststratification on quartiles of prior feedback received.

on the reply behaviors, but are less than an order of magnitude smaller. Furthermore, when egos receive more feedback, they also share more new posts during the experiment: A 10% increase in feedback causes a 0.7% increase in creating new posts.

We also computed estimates with other first-stage specifications: only the conversation salience factor, only the two encourage initiation factors, and a high-dimensional specification. Specifically, to potentially use heterogeneity in the true first-stage model, we fit a lasso (i.e., L1 penalized) first-stage model (53, 54) with both factors, interactions, and interactions with the stratum-defining variables, with the selected model having 23 nonzero coefficients. The results (Fig. 4) for feedback to peers and posting are statistically indistinguishable for all four models, whereas the two single-factor models differ for effects on the reply behaviors.

Comparison with Observational Estimates. In the absence of this peer encouragement design, scientists and decision makers could instead rely on observational data to study the effects of receiving feedback (10, 55). We thus evaluate how observational estimates compare with our experimental results. We regress each of the ego outcomes on observed feedback received, adjusting for strata and network clusters, as in the IV analysis, but ignoring assignment to peer encouragement conditions. (This analysis includes some variation in feedback received caused by the experiment, but this is a very small fraction of the variance, and it does not materially affect the results.) For all outcomes, these observational estimates of the effect of receiving feedback are substantially different from IV estimates from the peer encouragement design (Fig. 5). For the main outcomes of interest (posting and feedback to others), the observational coefficient estimates are 317–498% larger. On the other hand, they appear to underestimate reply behaviors by 36% and 68% for comments and likes, respectively. That is, in contrast to claims that observational estimates can upper bound true peer effects (6), the sign of the implied large-sample bias of the observational estimators varies across outcomes. These differences could be attributed to confounding, simultaneity, or the fact that IV and observational analyses often estimate different causal quantities (52).

Robustness to Dependence and Nuisance Interference. The preceding inferential results use a network adjacency- and cluster-robust estimator of the variance–covariance matrix (56, 57) to compute SEs; see *SI Appendix*, *Randomization Inference with Sensitivity Analysis*. To further examine the robustness of the results to nuisance interference, we used Fisherian randomization inference for the effect of feedback received on posting (which was the least statistically significant with $p = 0.0013$), while allowing for inference according to Assumption 3 with $c = 1$ under a model whereby an ego's outcome depends the assignments of their peers. This estimate remained statistically significant in the

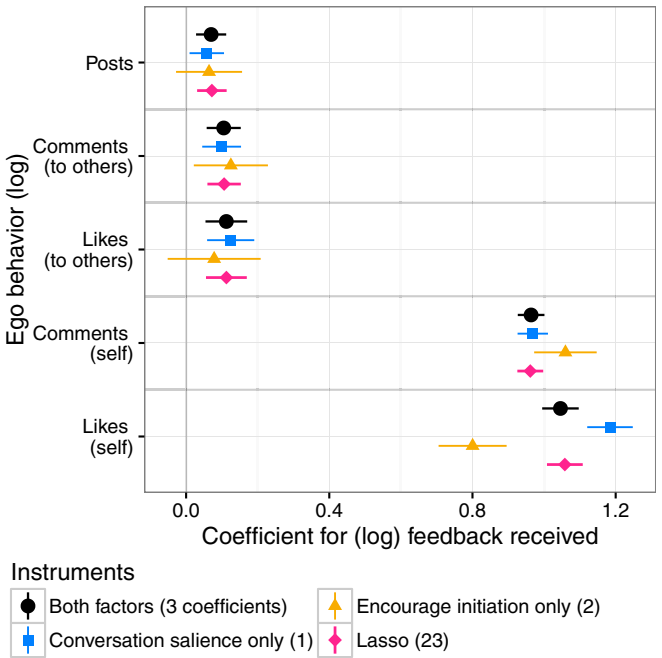

Fig. 4. Effects of receiving feedback on five ego behaviors, as estimated using IV analysis of the peer encouragement design with both peer encouragement factors (black), with only one of the factors (blue and yellow), and using a lasso (L1 penalized) first stage. Points are coefficient estimates from a log–log model. Comments (self) and likes (self) are the ego's comments on their own posts and likes of comments on their posts; these outcomes were expected to be most directly affected by receiving more feedback. The other ego behaviors involve giving more feedback to others and making new posts. Number of instruments is shown in parentheses; for the lasso, this counts only nonzero coefficients. Error bars are 95% network adjacency- and cluster-robust confidence intervals. *SI Appendix*, Table S2 displays these results in tabular form.

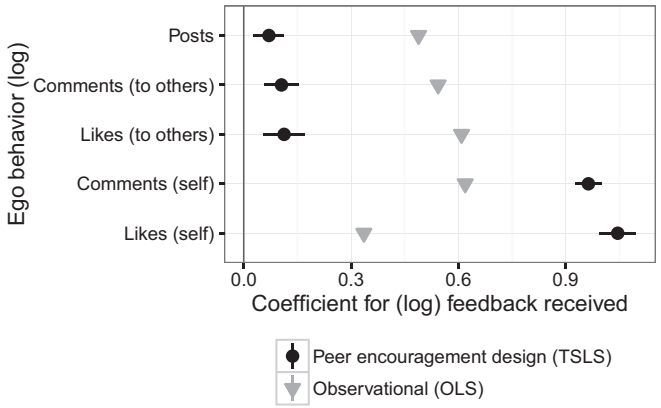

Fig. 5. Comparison of estimates using IV analysis of the peer encouragement design (black) and observational analysis using ordinary least squares (OLS) regression (gray). Points are coefficient estimates from a log–log model. Compared with the IV estimates, the observational analysis of ego behaviors on feedback received either substantially overestimates or underestimates peer effects. Error bars are 95% network adjacency- and cluster-robust confidence intervals and, for the observational (OLS) estimates, are smaller than the points.

presence of additive interference (maximum $p = 0.012$) or interactive interference (maximum $p = 0.017$) from peers; most of this difference in inference arises from the use of randomization inference with the rank sum test statistic, rather than allowing for interference per se (without interference, $p = 0.009$).

Discussion

Peer encouragement designs can be an effective strategy for estimating peer effects in networks: By randomly encouraging peers to specific behaviors, researchers can learn about the effects of those behaviors on egos. In this paper, we reviewed this class of experimental designs and demonstrated the potential to use a minimal encouragement (here, a small change to the user interface for giving feedback) to an ego-specific behavior. We found that receiving additional feedback causes individuals to give feedback to others and to share new posts. Compared with direct reply behaviors, these effects are smaller but still very substantial. This provides new evidence for the influence of peer effects in the use of communication technologies. It also informs our understanding of the value of social feedback to its recipients, as reflected in recipients' decisions to continue using a medium. In particular, receiving more feedback causes individuals to more frequently repeat the same behavior (posting content) that made them able to receive feedback in the first place. These results are informative about the role of directed behaviors in the adoption of technologies that enable both undirected (broadcast) and directed communications.

One limitation of this experiment is that it does not elucidate the mechanisms by which receiving feedback affects egos or distinguish different types of feedback. The observed effects are expected to occur for many reasons. For example, effects on giving feedback to others could be due to a psychological response (e.g., generalized reciprocity), or occur simply because receiving feedback causes users to return to Facebook more often, and therefore creates more opportunities to comment on peers' posts. Distinguishing these and other mechanisms would be difficult, but additional studies could test alternative explanations. For simplicity, we have focused on an experiment that identifies undifferentiated effects of feedback. Additional peer encouragement designs could also distinguish among different types of feedback.

A peer encouragement design identifies an encouragement-specific quantity: the effect of receiving additional feedback for egos' whose peers are induced by the encouragement to provide more feedback. This quantity is an ACR, the generalization of a local average treatment effect (LATE) for a multivalued treatment, or a weighted combination of ACRs. In this study, these are weighted average effects of feedback that would occur (or not) depending on small changes to the user interface. The conventional wisdom (cf. refs. 58–60) is that a LATE or ACR is less relevant than quantities that average over other, larger sets of potential outcomes, such as an average treatment effect or ATT. We argue that an ACR, in fact, averages over differences in peer behavior that are realistic under many relevant alternative policies. Researchers, marketers, or policy makers may be particularly interested in the average effects of incremental peer behaviors—behaviors that will occur or not depending on realistic changes to the environment, policy, or marketing campaign. In some standard economic models, the LATE is a piecewise constant approximation to this marginal treatment effect (61). Thus, if design, policy, or marketing decisions are expected to produce shifts in peer behaviors similar to the encouragement design, then the LATE or ACR may be of greater substantive relevance (cf. ref. 62). Of course, even different encouragements might define quite different ACRs. This experiment included two different peer encouragements that are expected to cause feedback at different times in the lifecycle of an ego's post: The encourage initiation factor should primarily cause the first feedback to occur at all or earlier, and the conversation salience factor should generate additional feedback on posts with existing feedback. We find that, despite this difference, these two factors identify quite similar ACRs, especially for the primary outcomes (*SI Appendix*, Fig. S8). This could increase our confidence that the present results may be informative about other attempts to cause people to receive more feedback on their posts. In this particular case, it is unclear what other averages would be preferable, because the natural generalization of the ATT to a multivalued variable like feedback received would average over contrasts comparing outcomes when egos receive their status quo levels of feedback and if they were to receive no feedback [i.e., $Y_i(D_i) - Y_i(0)$]. This thus includes contrasts in which very active, high-degree egos who currently receive large amounts of feedback receive none, which is perhaps unlikely to occur under policies being considered.

The current work highlights the advantages of large data sets and novel experimental designs for causal inference about how people affect each other. Our peer encouragement design provides credible causal estimates for the effects of receiving social feedback on Facebook; this is, to our knowledge, the first experimental evidence for these effects. The plausibility of a key assumption in our model, the exclusion restriction, partly depends on encouragements being minimal. However, encouragements that produce minimal variation can result in imprecise IV estimates; even studies with hundreds of thousands of observations will often suffer from the instruments being too weak (63). Peer encouragement designs with such minimal encouragements thus require a very large sample size and careful design (e.g., the ego-specific design used here) to estimate peer effects precisely. When feasible, however, peer encouragement designs can provide valuable insights into real-world social dynamics that can inform social science and policy decisions.

Materials and Methods

The peer encouragement design ran for 3 wk between September and October 2012. The egos in the data analyzed are 48.9 million Facebook users globally who had created at least one status update in the 4 d before the start of the experiment or during the experiment, who had at least one friend frequently using Facebook via the web interface, and who reported their age as at least 18. Note that 52% of these egos are randomly assigned to the peer encouragement condition reflecting the status quo at the time. Approximately 905 million Facebook users were peers of the egos and used the web interface. Further details about the sample and covariate balance are reported in *SI Appendix*, Table S1. The primary results we report are adjusted

with a set of sparse binary covariates (i.e., dummies) for quartiles of three pretreatment variables (forming $4^3 = 64$ strata) and 80,001 clusters formed by graph partitioning (see *SI Appendix*).

This study uses data from an experiment conducted for routine product improvement purposes and that posed no more than minimal risk. D.E. and E.B. designed and conducted the experiment as part of product development while employees of Facebook in 2012. Research using this data is consistent with the Data Policy that people accept when they choose to use the Facebook service. Accordingly, we did not separately notify users of this specific product test, nor did we obtain written informed consent. R.F.K. later contributed to this research using this existing data while an employee of Facebook in 2014 and 2015. Because he intended to use his university affiliation in reports on this study, R.F.K. asked the Stanford University institutional review board (IRB) to review a protocol for use of this previously collected anonymized data; the Stanford IRB approved this protocol. Similarly, when D.E. became a member of the Massachusetts Institute of Technology (MIT) faculty in 2015, the MIT IRB determined that a protocol for use of this previously collected anonymized data was exempt, and approved the protocol.

ACKNOWLEDGMENTS. This work benefited from comments by A. Fradkin, G. W. Imbens, S. Messing, A. Peysakhovich, J. Sekhon, S. J. Taylor, members of the Facebook Core Data Science team, and anonymous reviewers. We thank L. Backstrom, K. Deeter, and D. Vickrey for assistance conducting this experiment.

1. Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years. *N Engl J Med* 357(4):370–379.
2. Christakis NA, Fowler JH (2008) The collective dynamics of smoking in a large social network. *N Engl J Med* 358(21):2249–2258.
3. Cohen-Cole E, Fletcher JM (2008) Detecting implausible social network effects in acne, height, and headaches: Longitudinal analysis *BMJ* 337:a2533.
4. Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociol Methods Res* 40(2):211–239.
5. Thomas AC (2013) The social contagion hypothesis: Comment on 'Social contagion theory: Examining dynamic social networks and human behavior.' *Stat Med* 32(4): 581–590, and discussion (2013) 32(4):597–599.
6. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci USA* 106(51):21544–21549.
7. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annu Rev Sociol* 27:415–444.
8. Manski CF (2000) Economic analysis of social interactions. *J Econ Perspect* 14(3): 115–136.
9. Ogburn EL, van der Weele TJ (2014) Causal diagrams for interference. *Stat Sci* 29(4): 559–578.
10. Burke M, Marlow C, Lento T (2009) Feed me: Motivating newcomer contribution in social network sites. *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (Assoc Comput Machinery, New York), pp 945–954.
11. Christakis NA, Fowler JH (2013) Social contagion theory: Examining dynamic social networks and human behavior. *Stat Med* 32(4):556–577.
12. Ugander J, Backstrom L, Marlow C, Kleinberg J (2012) Structural diversity in social contagion. *Proc Natl Acad Sci USA* 109(16):5962–5966.
13. Jacobs AZ, Way SF, Ugander J, Clauset A (2015) Assembling thefacebook: Using heterogeneity to understand online social network assembly. arXiv:1503.06772.
14. Shriver SK, Nair HS, Hofstetter R (2013) Social ties and user-generated content: Evidence from an online social network. *Manage Sci* 59(6):1425–1443.
15. Tucker C (2008) Identifying formal and informal influence in technology adoption with network externalities. *Manage Sci* 54(12):2024–2038.
16. Moffitt RA (2001) Policy interventions, low-level equilibria, and social interactions. *Social Dynamics*, eds Durlauf SN, Young HP (MIT Press, Cambridge, MA), pp 45–82.
17. Walker D, Muchnik L (2014) Design of randomized experiments in networks. *Proc IEEE* 102(12):1940–1951.
18. Bakshy E, Eckles D, Yan R, Rosenn I (2012) Social influence in social advertising: Evidence from field experiments. *Proceedings of the 13th ACM Conference on Electronic Commerce* (Assoc Comput Machinery, New York), pp 146–161.
19. Asch SE (1956) Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol Monogr* 70(9):1–70.
20. Deutsch M, Gerard HB (1955) A study of normative and informational social influences upon individual judgment. *J Abnorm Soc Psychol* 51(3):629–636.
21. Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.
22. Sacerdote B (2001) Peer effects with random assignment: Results for Dartmouth roommates. *Q J Econ* 116(2):681–704.
23. Carrell SE, Fullerton RL, West JE (2009) Does your cohort matter? Measuring peer effects in college achievement. *J Labor Econ* 27(3):439–464.
24. Kling JR, Liebman JB, Katz LF (2007) Experimental analysis of neighborhood effects. *Econometrica* 75(1):83–119.
25. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197.
26. Aral S, Walker D (2011) Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Manage Sci* 57(9):1623–1639.
27. Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. *Science* 337(6092):337–341.
28. Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. *Proceedings of the 21st International Conference on World Wide Web* (Assoc Comput Machinery, New York), pp 519–528.
29. Pearl J (2009) Causal inference in statistics: An overview. *Stat Surv* 3:96–146.
30. Holland PW (1988) Causal inference and path analysis. *Sociol Methodol* 18:449–484.
31. Bogatz GA, Ball S (1971) *The Second Year of Sesame Street: A Continuing Evaluation: A Report to the Children's Television Workshop* (Educ Test Serv, Princeton), Vol 1.
32. Powers DE, Swinton SS (1984) Effects of self-study for coachable test item types. *J Educ Psychol* 76(2):266–278.
33. Angrist J, Imbens G (1995) Two-stage lease squares estimation of average causal effects in models with variable treatment intensity. *J Am Stat Assoc* 90(430):431–442.
34. Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. *Econometrica* 62(2):467–475.
35. Angrist JD, Imbens GW, Rubin DB (1996) Identification of causal effects using instrumental variables. *J Am Stat Assoc* 91(434):444–455.
36. Imbens GW (2014) Instrumental variables: An econometrician's perspective. *Stat Sci* 29(3):323–358.
37. Imbens GW, Rubin DB (2015) *Causal Inference in Statistics, Social, and Biomedical Sciences* (Cambridge Univ Press, Cambridge, UK).
38. Manski CF (1993) Identification of endogenous social effects: The reflection problem. *Rev Econ Stud* 60(3):531–542.
39. Athey S, Eckles D, Imbens GW (2015) Exact p-values for network interference. arXiv: 1506:02084.
40. Angelucci M, De Giorgi G (2009) Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *Am Econ Rev* 99(1):486–508.
41. Duflo E, Saez E (2003) The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *Q J Econ* 118(3):815–842.
42. Forastiere L, Mealli F, VanderWeele TJ (December 22, 2015) Identification and estimation of causal mechanisms in clustered encouragement designs: Disentangling bed nets using Bayesian principal stratification. *J Am Stat Assoc*, 10.1080/01621459.2015.1125788.
43. Miguel E, Kremer M (2004) Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72(1):159–217.
44. Nickerson DW (2008) Is voting contagious? Evidence from two field experiments. *Am Polit Sci Rev* 102(1):49–57.
45. Bond RM, et al. (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415):295–298.
46. Cai J, De Janvry A, Sadoulet E (2012) Social networks and the decision to insure. *Am Econ J Appl Econ* 7(2):81–108.
47. Coppock A, Guess A, Ternovski J (2016) When treatments are tweets: A network mobilization experiment over Twitter. *Polit Behav* 38(1):105–128.
48. Frey BS, Jegen R (2001) Motivation crowding theory. *J Econ Surv* 15(5):589–611.
49. Cox DR (1958) *Planning of Experiments* (Wiley, New York).
50. Manski CF (2013) Identification of treatment response with social interactions. *Econom J* 16(1):S1–S23.
51. Aronow P, Samii C (2014) Estimating average causal effects under general interference. Working paper (Washington Univ St. Louis). Available at polmeth.wustl.edu/node/233.
52. Lochner L, Moretti E (2015) Estimating and testing models with many treatment levels and limited instruments. *Rev Econ Stat* 97(2):387–397.
53. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc B* 58(1):267–288.
54. Belloni A, Chen D, Chernozhukov V, Hansen C (2012) Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80(6): 2369–2429.
55. Burke M, Kraut RE (2014) Growing closer on Facebook: Changes in tie strength through social network site use. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Assoc Comput Machinery, New York), pp 4187–4196.
56. Conley TG (1999) GMM estimation with cross sectional dependence. *J Econom* 92(1): 1–45.
57. Cameron AC, Gelbach JB, Miller DL (2011) Robust inference with multiway clustering. *J Bus Econ Stat* 29(2):238–249.
58. Aronow PM, Carnegie A (2013) Beyond LATE: Estimation of the average treatment effect with an instrumental variable. *Polit Anal* 21(4):492–506.
59. Deaton A (2010) Instruments, randomization, and learning about development. *J Econ Lit* 48(2):424–455.
60. Imbens GW (2010) Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *J Econ Lit* 48(2):399–423.
61. Heckman JJ, Vytlacil EJ (1999) Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proc Natl Acad Sci USA* 96(8): 4730–4734.
62. Dunning T (2012) *Natural Experiments in the Social Sciences: A Design-Based Approach* (Cambridge Univ Press, Cambridge, UK).
63. Bound J, Jaeger DA, Baker RM (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J Am Stat Assoc* 90(430):443–450.