This paper examines a commonly used measure of persuasion whose precise interpretation has been obscure in the literature. By using the potential outcome framework, we define the causal persuasion rate by a proper conditional probability of taking the action of interest with a persuasive message conditional on not taking the action without the message. We then formally study identification under empirically relevant data scenarios and show that the commonly adopted measure generally does not estimate—but often overstates—the causal rate of persuasion. We discuss several new parameters of interest and provide practical methods for causal inference.
I. Introduction

How effectively one can persuade one’s audience has been of interest to ancient Greek philosophers in the Lyceum of Athens, early–modern English preachers in St Paul’s Cathedral, and contemporary American news producers at Fox News in New York City. Recently, economists have been endeavoring to build theoretical models of persuasion (e.g., Kamenica and Gentzkow 2011; Che, Dessein, and Kartik 2013; Gentzkow and Kamenica 2017; Prat 2018; Bergemann and Morris 2019) and to quantify empirically the extent to which persuasive efforts affect the behavior of consumers, voters, donors, and investors (for a survey of the recent literature, see DellaVigna and Gentzkow 2010).

Since DellaVigna and Kaplan (2007) proposed a measure of persuasion (hereafter, DK measure), it has been used and modified by many authors (e.g., Enikolopov, Petrova, and Zhuravskaya 2011; Gentzkow, Shapiro, and Sinkinso 2011; DellaVigna et al. 2014; Bassi and Rasul 2017; Martin and Yurukoglu 2017; Chen and Yang 2019) to quantify the persuasive effects of informational treatment. However, its precise interpretation has been obscure because of a lack of formal identification analysis. In fact, we show that the commonly used measure of persuasion does not estimate the causal rate of persuasion on any subpopulation in general. Therefore, it is misleading to call the DK measure the persuasion rate, although it is a common practice in the literature; instead, we will reserve the term of the persuasion rate for the population parameter that is properly defined by a conditional probability in the potential outcome framework.

The flaw of the DK measure arises from failing to distinguish a local average treatment effect (LATE; see Imbens and Angrist 1994) from the average treatment effect (ATE), where the difference between the two can be substantial for a heterogeneous population; specifically, the DK measure rescales a LATE with a factor that is relevant only for the ATE. For instance, in DellaVigna and Kaplan’s (2007) example, even if Fox News has a high persuasive effect among those who will watch the channel if and only if it is available through a local cable package, it may not be persuasive at all when other people—such as Democrats or Democrat-leaning independents—are all included.

Focusing on the case of binary outcomes, we analyze the problem of measuring persuasive effects of informational treatment through the lens of the potential outcome framework. Specifically, we define the persuasion rate by a proper conditional probability of the agent taking an action of interest with a persuasive message, given that the agent does not take

1 See Rapp (2010) for three technical means of persuasion in Aristotle’s Rhetoric.
2 See Kirby (2008) for historic details of the public persuasion at Paul’s Cross, the open-air pulpit in St Paul’s Cathedral in the sixteenth century.
3 DellaVigna and Kaplan (2007) and Martin and Yurukoglu (2017) measure the persuasive effects of slanted news, using data on Fox News.
the action without the conveyed message. We then formally study identification under a few empirically relevant scenarios of data availability. Our analysis will articulate what the DK measure of persuasion estimates, why it is misleading, and how we can fix the problem.

While the persuasion rate is concerned with the entire population (and hence related to the ATE), we also consider a local persuasion rate that focuses on the group of compliers. Our identification analysis shows that the DK measure estimates neither a local nor an average persuasion rate.

Our identification analyses are based on a few empirically relevant scenarios of data availability. We do this because the problem of data availability is particularly important in the context of measuring persuasive effects. For instance, individual-level partisan vote outcome data rarely exist because of confidentiality issues. Thus, the outcome variable is frequently measured only at an aggregate level. Also, it is not always the case to observe an actual exposure to a persuasive message in the same dataset along with the outcome and instrument. Indeed, DellaVigna and Kaplan’s (2007) analysis uses a micro dataset for the treatment and instrument and a separate aggregate dataset for the outcome and instrument. In order to address these challenges, we consider three different scenarios of data availability explicitly: given the instrument, (i) the outcome and treatment are jointly observed, (ii) they are observed separately, or (iii) the treatment is not observed at all. We obtain the sharp bounds on the persuasion rate for the entire population as well as other subpopulations of potential interest. Therefore, our work builds on the econometrics literature on partial identification (e.g., Manski 2003, 2007; Tamer 2010) as well as the literature on program evaluation (for surveys of the literature, see Heckman and Vytlacil 2007a; Imbens and Wooldridge 2009).

The main findings of this paper can be summarized as follows. If there is no heterogeneity in the population, then the DK measure of persuasion estimates the rate of persuasion, provided that a simple monotonicity assumption is imposed; however, this case is an exception rather than the rule. Indeed, the rate of persuasion is only partially identified as an interval in general, where the sharp lower bound generally corresponds to the DK measure multiplied by the relative size of the complier group regardless of the specific data scenario. Therefore, the DK measure is strictly larger than the lower bound whenever there is partial compliance. It is also remarkable that the data scenarios matter only for the sharp upper bound. Therefore, the value of jointly observing the treatment and outcome only lies in obtaining a potentially more informative upper bound on the persuasion rate. We also investigate identification of the local persuasion rate (i.e., the persuasion rate for the group of compliers) under the same three data scenarios. It is point identified under the most favorable data scenario but only partially identified under the other scenarios; even in the case of point identification, the DK measure generally
differs from the local persuasion rate. If a continuous instrument is available, then we can target a *marginal persuasion rate* that is akin to the marginal treatment effect (e.g., Heckman and Vytlacil 2005). Therefore, having a continuous instrument opens up the possibility of point identification of the persuasion rate for a policy-relevant population if the instrument is sufficiently rich.

In order to illustrate our findings, we discuss two empirical examples in the main text, while we provide a few more in the appendixes. First, we revisit Chen and Yang (2019), where the interest is in the effect of Chinese students having access to uncensored media on behaviors, beliefs, and attitudes; we use the same variables and setup as in the original paper for this exercise. Second, we analyze the voting behavior and newspaper readership, using the data from Gerber, Karlan, and Bergan (2009). Overall, we show that the DK measure of persuasion tends to overstate the persuasive effects while masking underlying heterogeneity.

The remainder of the paper is organized as follows. In section II, we recall some backgrounds and specifics of the DK measure of persuasion. Here, we properly define the persuasion rate at the population level by using the potential outcome framework. In section III, we discuss identification of the persuasion rate. In section IV, we study the local and marginal versions of the persuasion rate. In section V, we discuss our recommendations on what to do in practice, including practical inferential issues, and clarify the difference between a population version of the DK measure of persuasion and the persuasion rate. In section VI, we provide two empirical illustrations. We conclude in section VII. The appendixes include additional results and examples, including an extension to nonbinary outcomes, a detailed discussion about methods for inference, and all of the proofs.

II. Background

It is helpful to recall DellaVigna and Kaplan (2007) as a prototypical example, where they study the effect of an exposure to Fox News on the probability of voting for a Republican presidential candidate. Here, the informational treatment of interest is the viewership of the Fox News channel, and the persuasion rate of the media can be thought of as the proportion of the Fox News viewers who voted for a Republican candidate among those who would not have done so if they had not watched Fox News. It should be noted that the agents’ decisions about whether to watch Fox News may

---

Stata commands for estimation and inference based on this paper’s identification results are publicly available at `https://github.com/persuasio`. Alternatively, they can be installed from within Stata by typing "`ssc install persuasio.`"
be correlated with their political orientation. In order to address the endogeneity issue, DellaVigna and Kaplan (2007) use Fox News availability via local cable in the year of 2000 as an instrumental variable.

In efforts to measure the persuasion rate as explained above, DellaVigna and Kaplan (2007) and DellaVigna and Gentzkow (2010) propose the (infeasible) estimand $f$, defined as follows: for a binary outcome and by using DellaVigna and Kaplan’s (2007) notation

$$f = \frac{y_T - y_C}{e_T - e_C} \cdot \frac{1}{1 - y_0},$$

where $T$ and $C$ represent an instrument assignment status, such as having Fox News available via local cable. Here, for $j \in \{T, C\}$, $y_j$ is the share of group $j$ taking the action of interest (e.g., voting for a Republican candidate), and $e_j$ is the share of group $j$ exposed to a persuasive message. Further, $y_j$ is the share of those who would take the action of interest without listening to the persuasive message. So, $f$ is a rescaled version of the usual Wald statistic that estimates the LATE, where rescaling is apparently to obtain a rate that focuses on those who are to be persuaded. As DellaVigna and Kaplan (2007) noted, $y_0$ involves a counterfactual that is often unobserved. For this reason, $f$ is generally not a feasible estimand, and DellaVigna and Kaplan (2007) propose using $y_C$ in place of $y_0$ as an approximation, which yields a feasible estimand, say $\tilde{f}$. We will refer to $f$ or its feasible approximation $\tilde{f}$ by the DK measure of the persuasion rate.

Without a formal justification, it is common in the literature to interpret $f$ as a conditional probability. For instance, DellaVigna and Kaplan (2007, 1218) explain, “The key parameter is $f$, the fraction of the audience that is convinced by Fox News to vote Republican.” Similar interpretations are prevalent in the literature, as the following quotations demonstrate:

Whenever possible, we report the results in terms of the persuasion rate (DellaVigna and Kaplan 2007), which estimates the percentage of receivers that change the behavior among those that receive a message and are not already persuaded. (DellaVigna and Gentzkow 2010, 645)

We can also translate our estimates into a “persuasion rate” (DellaVigna and Kaplan 2007)—the number of eligible voters who changed their voting behavior as a result of the introduction of the newspaper, as a fraction of all those who could have changed their behavior. (Gentzkow, Shapiro, and Sinkinson 2011, 3003)

Therefore, $T$ and $C$, which appear to denote treatment and control groups, should be understood as the status of the intent to treat (ITT), not the actual treatment status.
The persuasion rate is the fraction of the audience of a media outlet who are convinced to change their behavior (in this case, their vote) as a result of being exposed to this media outlet. (DellaVigna et al. 2014, 125)

Finally, we computed estimates of DellaVigna and Kaplan’s (2007) concept of persuasion rates: the success rate of the channels at converting votes from one party to the other. The numerator in the persuasion rate here is the number of, for example, the Fox News Channel (FNC) viewers who are initially Democrats but by the end of an election cycle change to supporting the Republican party. The denominator is the number of FNC viewers who are initially Democrats. (Martin and Yurukoglu 2017, 2590)

Also, in their survey, DellaVigna and Gentzkow (2010) use \( \tilde{f} \)— that is, a feasible approximation of \( f \)— as a key summary statistic to compare persuasive effects across different studies.

However, as we mentioned in the introduction, neither \( f \) nor \( \tilde{f} \) estimates the persuasion rate—that is, the intended conditional probability—on any subpopulation in general. Therefore, it is misleading to label \( f \) or \( \tilde{f} \) as a persuasion rate, and it is generally invalid to make comparisons across different studies. Indeed, \( f \) or \( \tilde{f} \) may not even be a proper conditional probability in a heterogeneous population: for example, the approximation \( \tilde{f} \) can even be larger than 1. We will articulate under what assumptions \( \tilde{f} \) turns out to be a proper conditional probability and how we should interpret it in its relationship with the persuasion rate.

In order to facilitate our discussion, we start with formally defining the persuasion rate at the population level by using the potential outcome framework. Let \( T_i \) denote the binary indicator that equals 1 if individual \( i \) is exposed to persuasive information such as Fox News. Let \( Y_i(t) \) be a binary indicator that shows agent \( i \)'s potential action when \( T_i \) is set to \( t \in \{0, 1\} \). For example, \( Y_i(1) \) equals 1 if individual \( i \) votes for a Republican candidate after watching Fox News. The econometrician never observes both \( Y_i(0) \) and \( Y_i(1) \) but can observe only either of the two: that is, \( Y_i = T_iY_i(1) + (1 - T_i)Y_i(0) \).

Then, the fraction of the people who take the action of interest with an exposure to a persuasive message among those who would not without it is given by

\[
\theta_{pr} = \mathbb{P}\{Y_i(1) = 1 \mid Y_i(0) = 0\},
\]

provided that the conditional probability is well defined: \( \theta_{pr} \) is the persuasion rate at the population level. Using conditional probability is to rule

\[6\] So, both \( Y_i \) and \( T_i \) are binary. In app. B, we extend our results to the case where the potential outcomes are multinomial.
out the case of “preaching to the converted”; if $Y_i(0) = 1$, then those individuals are already persuaded to take the action of interest even without the persuasive treatment, and therefore we do not count them in defining the persuasion rate.$^7$

The common estimand $\tilde{f}$ (or even $f$) does not estimate $\theta_{pr}$ in a heterogeneous population; in fact, even in a homogeneous population, $\tilde{f}$ or $f$ cannot be (asymptotically) equated with $\theta_{pr}$ without an extra monotonicity assumption. However, the rescaled quantity $(e_T - e_C)\tilde{f}$, which is always no greater than $\tilde{f}$, does provide valid information about $\theta_{pr}$ in that it corresponds to the sharp lower bound of the identified interval of $\theta_{pr}$ in general. The best way to clarify all the issues is to conduct a rigorous analysis on the identification of $\theta_{pr}$, which is our next topic. For quick takeaways, see section V.

### III. Identification of the Persuasion Rate

Identification of $\theta_{pr}$ is challenging for various reasons, including the fact that $\theta_{pr}$ depends on the joint distribution of the potential outcomes and that $T_i$ can be endogenous, and it is often difficult to observe $T_i$ and $Y_i$ jointly. To allow for potential endogeneity, we use a binary instrument, $Z_i$, throughout the paper unless otherwise specified. Exogenous covariates, $X_i$, may be observed, but we suppress $X_i$ from our identification analysis. In other words, we implicitly assume throughout the paper that all assumptions and results are conditional on the value of $X_i$. Therefore, the observed variables are $Y_i$, $T_i$, $Z_i$, all of which are binary throughout the main text. See appendix D for how to deal with $X_i$ in practice. Also, see appendix B for an extension to the case of multinomial outcomes.

The data issue on $T_i$ is addressed by considering three scenarios of data availability. Specifically, for the purpose of the identification analysis, we assume that for $(y, t, z) \in \{0, 1\}^3$, (i) $P(Y_i = y, T_i = t \mid Z_i = z)$ is known, (ii) $P(Y_i = y \mid Z_i = z)$ and $P(T_i = t \mid Z_i = z)$ are separately known, or (iii) $P(Y_i = y \mid Z_i = z)$ is all that is known, depending on the specific scenario of interest. For example, DellaVigna and Kaplan (2007) use town-level election data to estimate $P(Y_i = y \mid Z_i = z)$ and micro-level media audience data to infer $P(T_i = t \mid Z_i = z)$, which corresponds to case ii.$^8$

It requires an additional assumption to address the challenge that $Y_i(1)$ and $Y_i(0)$ are never simultaneously observed while $\theta_{pr}$ depends on their

---

$^7$ The idea of using conditional probability to define a parameter of interest can also be found in Heckman, Smith, and Clements (1997), though their context is quite different from ours.

$^8$ Throughout the discussion, we assume that $T_i$ is correctly measured if it is observed. See Ura (2018), Nguimkeu, Denteh, and Tchernis (2019), and Calvi, Lewbel, and Tommasi (2022) for the issues of mismeasured treatment. Their subject matter is distinct from ours.
joint distribution. Before we present our identification results, we discuss our key assumptions in section III.A.

A. The Key Assumptions

Our first key assumption is that the persuasive message is directional, which will be important to decouple $\theta_{pr}$ by the marginals of the potential outcomes.

**Assumption A (Monotonic treatment response).** The potential outcomes $Y_i(1)$ and $Y_i(0)$ are binary, and they satisfy $Y_i(0) \leq Y_i(1)$ with probability 1.

Assumption A is a binary version of the monotonic treatment response assumption used in Manski (1997) and Manski and Pepper (2000). Assumption A allows $Y_i(0) = Y_i(1)$ with probability 1, and therefore it does not rule out the possibility that watching Fox News has no impact on the agent’s behavior at all. The inequality in assumption A means that the messages Fox News delivers are biased or directional in favor of Republican candidates; that is, if a voter is going to vote for a Republican candidate without watching Fox News, then watching Fox News will not change that. In other words, assumption A rules out the possibility that the level of distrust that a voter has on Fox News is so high that she takes actions based on the opposite of the messages Fox News delivers.

Since assumption A is a key assumption in the paper, we first clarify how much we can hope for with and without assumption A.

**Lemma 1.** We generally have

$$\max \left[ 0, \frac{\mathbb{P}\{Y_i(1) = 1\} - \mathbb{P}\{Y_i(0) = 1\}}{1 - \mathbb{P}\{Y_i(0) = 1\}} \right] \leq \theta_{pr} \leq \min \left[ \frac{\mathbb{P}\{Y_i(1) = 1\}}{1 - \mathbb{P}\{Y_i(0) = 1\}}, 1 \right],$$

(3)

where the bounds are sharp in that $\theta_{pr}$ can be anything between the bounds without changing the marginals $\mathbb{P}\{Y_i(1) = 1\}$ and $\mathbb{P}\{Y_i(0) = 1\}$. Further, assumption A holds if and only if $\theta_{pr} = \theta_{avg}$, where

$$\theta_{avg} := \frac{\mathbb{P}\{Y_i(1) = 1\} - \mathbb{P}\{Y_i(0) = 1\}}{1 - \mathbb{P}\{Y_i(0) = 1\}}.$$

(4)

Equation (3) is a consequence of the Fréchet-Hoeffding inequality on the probability of a joint event. Since the two potential outcomes are never observed simultaneously, all we can ever hope to identify is their marginals, and equation (3) expresses the (sharp) bounds on $\theta_{pr}$ in terms of the marginal probabilities of the potential outcomes.

---

9 In app. A, we present a simple economic model that motivates assumption A.
Suppose that there are some voters who have such a high level of distrust of Fox News that not watching Fox News would help them to take favorable actions to a Republican candidate but that those voters are only a minority and on average we still have a strict stochastic dominance relationship between $Y_i(1)$ and $Y_i(0)$ (i.e., $\mathbb{P}\{Y_i(1) = 1\} > \mathbb{P}\{Y_i(0) = 1\}$). Then, the lower bound on $\theta_{pr}$ will be ensured to be nontrivial.

Assumption A is stronger than the stochastic dominance, but it delivers a stronger result. In fact, assumption A is necessary and sufficient to express $\theta_{pr}$ in terms of the marginal probabilities of the counterfactual outcomes. In this case, the conditional probability $\theta_{pr}$ is the ATE divided by $\mathbb{P}\{Y_i(0) = 0\}$. Throughout the rest of the paper, we present most of our results by using assumption A not only because it is convenient but also because being biased or directional seems to be the nature of persuasive effort. However, we emphasize that $\theta_{avg}$, the rescaled version of the ATE, is always a valid lower bound on $\theta_{pr}$, as equation (3) shows.

The next assumption is concerned with the treatment assignment $T_i$ and the instrument $Z_i$.

**Assumption B (No defiers and an exogenous instrument).** The binary treatment $T_i$ has a threshold structure, that is,

$$T_i = \mathbb{1}\{V_i \leq e(Z_i)\}, \tag{5}$$

where $V_i$ is an unobserved random variable that is uniformly distributed on $[0, 1]$. The binary instrument $Z_i$ is independent of $(Y_i(t), V_i)$ for $t = 0, 1$. Finally, we have $0 \leq e(0) < e(1) \leq 1$.

Assumption B is standard for causal inference using instrumental variables. The ITT variable, $Z_i$ is randomly assigned; however, $T_i$ can be endogenous via the dependence between $V_i$ and $Y_i(t)$. The function $e(\cdot)$ is the propensity score or, more descriptively in our context, it can be referred to as the exposure rate.

If $T_i(z)$ denotes counterfactual treatment when a binary instrument takes value $z$, then agent $i$ is called a complier when $T_i(1) = 1$ and $T_i(0) = 0$; never-takers ($T_i(z) = 0$ for all $z$), always-takers ($T_i(z) = 1$ for all $z$), and defiers ($T_i(1) = 0$, $T_i(0) = 1$) are similarly understood. Under assumption B, $i$ is a complier if and only if $e(0) < V_i \leq e(1)$. Similarly, $i$ is an always-taker when $V_i \leq e(0)$, and she is a never-taker when $V_i > e(1)$. Therefore, under assumption B, there are only three groups in the population: that is, always-takers, never-takers and compliers. Indeed, as Vytlacil (2002) has shown, the threshold structure in equation (5) is equivalent to assuming the absence of defiers, which is a popular assumption in econometrics to identify the LATE.

In sections III.B and III.C, we present identification results for $\theta_{pr}$, which is the same as $\theta_{avg}$ under assumption A. Section III.B covers the simplest
case, where everybody complies so that there is no difference between the actual treatment and the ITT, that is, $T_i = Z_i$ for each $i$; we refer to this case as the sharp persuasion design, where there is no distinction among different data scenarios. In this case, not surprisingly, $\theta_{\text{avg}}$ is point identified from the distribution of $Y_i$ given $Z_i$. However, when $T_i$ and $Z_i$ are different, which we call the fuzzy persuasion design, we have only partial identification of $\theta_{\text{avg}}$, where each of the three data scenarios becomes relevant. Before we move on, we define

$$
\theta_L := \frac{\mathbb{P}(Y_i = 1 \mid Z_i = 1) - \mathbb{P}(Y_i = 1 \mid Z_i = 0)}{1 - \mathbb{P}(Y_i = 1 \mid Z_i = 0)},
$$

provided that $\mathbb{P}(Y_i = 1 \mid Z_i = 0) < 1$: $\theta_L$ is an identified parameter from the distribution of $Y_i$ given $Z_i$. It turns out that first, $\theta_L$ is equal to $\theta_{\text{avg}}$ under the sharp persuasion design, and second, it is the sharp lower bound on $\theta_{\text{avg}}$ in the fuzzy persuasion design regardless of which of the three data scenarios applies.

### B. The Sharp Persuasion Design

**Theorem 1.** Suppose that assumptions A and B hold. If $e(1) - e(0) = 1$ (i.e., $T_i = Z_i$ with probability 1), then for $z \in \{0, 1\}$, we have $\mathbb{P}\{Y_i(z) = 1\} = \mathbb{P}(Y_i = 1 \mid Z_i = z)$, and hence $\theta_{\text{pr}} = \theta_{\text{avg}} = \theta_L$.

The condition of $e(1) - e(0) = 1$ means that everybody is a complier, and hence there is essentially no difference between $T_i$ and $Z_i$; thus, the sharp design is equivalent to a situation where $T_i$ is observed and randomized. However, this is rather an exceptional situation in social sciences. The key identification question should be how far we can go when the design is not sharp (i.e., not everybody is a complier). We answer this question in section III.C.

Without assumption A, the general sharp identified bounds on $\theta_{\text{pr}}$ are given by

$$
\max\{0, \theta_L\} \leq \theta_{\text{pr}} \leq \min\left\{\frac{\mathbb{P}(Y_i = 1 \mid Z_i = 1)}{1 - \mathbb{P}(Y_i = 1 \mid Z_i = 0)}, 1\right\}.
$$

Therefore, even in the sharp persuasion design, $\theta_{\text{pr}}$ is only partially identified, and its sharp bounds can be trivial without assumption A. Even so, it is worth noting that $\theta_L$ remains a valid lower bound.

### C. The Fuzzy Persuasion Design

In the fuzzy design, the three scenarios of data availability we mentioned earlier become pertinent.
1. Identification with the Joint Distribution of \((Y_i, T_i, Z_i)\)

Even the full joint distribution of \((Y_i, T_i, Z_i)\) does not point identify the ATE. Recall that those with \(Z_i = 0\) and \(T_i = 0\) comprise the compliers and the never-takers, while those with \(Z_i = 0\) and \(T_i = 1\) are the always-takers. Similarly, those with \(Z_i = 1\) and \(T_i = 0\) are the never-takers, while those with \(Z_i = 1\) and \(T_i = 1\) consist of the compliers and the always-takers. Therefore, these four cases correspond to different subpopulations, and the only subpopulation that we can study for both \(T_i = 0\) and \(T_i = 1\) in common is that of compliers, which explains why the Wald statistic estimates the LATE, not ATE. For the same reason, \(\theta_{avg}\) cannot be point identified; however, we can derive its sharp bounds.

**Assumption C (Full observability).** The joint distribution of \((Y_i, T_i, Z_i)\) is known, where \(\Pr(Y_i = 0 \mid Z_i = 0) > 0\).

**Theorem 2.** Suppose that assumptions A–C are satisfied. Then, the sharp identified interval of \(\theta_{pr} = \theta_{avg}\) is given by \([\theta_L, \theta_U]\), where \(\theta_L\) is given in equation (6) and

\[
\theta_U := \frac{\Pr(Y_i = 1, T_i = 1 \mid Z_i = 1) - \Pr(Y_i = 1, T_i = 0 \mid Z_i = 0) + 1 - e(1)}{1 - \Pr(Y_i = 1, T_i = 0 \mid Z_i = 0)}.
\]

To prove theorem 2, we first derive the sharp identified bounds for \(\Pr\{Y_i(1) = 1\}\) and \(\Pr\{Y_i(0) = 1\}\) separately; we denote them by the intervals \([m_a, M_a]\) and \([m_b, M_b]\), respectively. These bounds are special cases of Manski and Pepper (2000) under the monotonic treatment response assumption coupled with the exogeneity of the instrument. Then, letting 

\[
a := \Pr\{Y_i(1) = 1\} \quad \text{and} \quad b := \Pr\{Y_i(0) = 1\},
\]

we obtain the upper bound of the identified interval of \(\theta_{avg}\) by solving

\[
\max_{a,b} \frac{a - b}{1 - b} \quad \text{subject to} \quad a \in [m_a, M_a], \quad b \in [m_b, M_b], \quad a \geq b,
\]

while the lower bound can be found by doing minimization instead of maximization. We then appeal to continuity and the intermediate value theorem for the sharpness result.

It is proved in appendix I that \(m_a = \Pr(Y_i = 1 \mid Z_i = 1)\) and \(M_b = \Pr(Y_i = 1 \mid Z_i = 0)\). Therefore, an examination of (8) reveals that the lower bound is attained when \(a = m_a\) and \(b = M_b\). To develop intuition behind theorem 2, we discuss what \(a = m_a\) and \(b = M_b\) means, for which the behavior of the never-takers and that of the always-takers matter. Since \(m_a = \Pr\{Y_i(1) = 1, T_i = 1 \mid Z_i = 1\} + \Pr\{Y_i(0) = 1, T_i = 0 \mid Z_i = 1\}\), we know that \(a = m_a\) holds when \(\Pr\{Y_i(0) = 1, T_i = 0 \mid Z_i = 1\} = \Pr\{Y_i(1) = 1, T_i = 0 \mid Z_i = 1\}\). Here, the event \(Z_i = 1, T_i = 0\) means that \(i\) is a never-taker, because defiers are assumed to be nonexistent. Therefore, \(a = m_a\) means that the treatment has no effect on the group of never-takers, unless there are no never-takers at all. Similarly, \(b = M_b\)
holds when \( \mathbb{P}\{Y(1) = 1, T_i = 1 \mid Z_i = 0\} = \mathbb{P}\{Y(0) = 1, T_i = 1 \mid Z_i = 0\} \).

Since \( Z_i = 0, T_i = 1 \) means that \( i \) is an always-taker, we know that \( b = M_b \) holds when the treatment does not affect the behavior of the always-takers, unless there are no always-takers at all. Therefore, \( \theta_{avg} \), which is the same as the persuasion rate \( \theta_{pr} \) under assumption A, is smallest when there are null treatment effects for both the never-takers and the always-takers.

Intuition for the upper bound can also be obtained by considering the noncomplier groups. The upper bound corresponds to the case where \( a = M_a \) and \( b = m_b \), where it is shown in appendix I that \( M_a = \mathbb{P}(Y_i = 1, T_i = 1 \mid Z_i = 1) + 1 - e(1) \) and \( m_b = \mathbb{P}(Y_i = 1, T_i = 0 \mid Z_i = 0) \).

Here, note that \( a = M_a \) is equivalent to \( \mathbb{P}\{Y(1) = 0, T_i = 0 \mid Z_i = 1\} = 0 \) and \( b = m_b \) is equivalent to \( \mathbb{P}\{Y(0) = 1, T_i = 1 \mid Z_i = 0\} = 0 \). Therefore, we can see that the persuasion rate \( \theta_{avg} \) equals the upper bound when every never-taker has \( Y_i(1) = 1 \) and none of the always-takers have \( Y_i(0) = 1 \); for example, all those who never watch Fox News (whether it is available or not) would actually have voted for a Republican candidate if they had watched it, and all those who always watch Fox News would not have voted for a Republican without watching the channel.

The bounds in theorem 2 shrink to a singleton as \((e(0), e(1))\) approaches \((0, 1)\), which is consistent with the result in theorem 1. Also, it is worth noting that the lower bound \( \theta_L \) depends on only the distribution of \((Y_i, Z_i)\): observing \( T_i \) along with \((Y_i, Z_i)\) helps only for the upper bound. If \( e(1) \) is too small, then the upper bound will not be very informative: \( \theta_U \) converges to 1 as \( e(1) \) approaches 0; that is, if nobody reads a newspaper when they receive free subscriptions, then we do not learn much about how persuading the newspaper is. However, even if \( e(1) \) approaches 1, the upper bound does not necessarily shrink to the lower bound; for example, we do not necessarily pin down the persuasion rate of reading the newspaper even if everybody who has free subscriptions actually reads it.

We now establish partial identification of \( \theta_{pr} \) without assumption A, that is, \( \theta_{pr} \neq \theta_{avg} \), for the sake of completeness. Let \( NT = \{V_i > e(1)\} \) and \( AT = \{V_i \leq e(0)\} \) be the event of \( i \) being a never-taker and an always-taker, respectively.

**Theorem 3.** Suppose that assumptions B and C are satisfied.

1. If \( \mathbb{P}(Y_i = 0 \mid Z_i = z) > 0 \) for \( z = 0, 1 \), then the sharp identified interval of \( \theta_{pr} \) is given by

\[
\max \left\{ 0, \frac{\mathbb{P}(Y_i = 1, T_i = 1 \mid Z_i = 1) - \mathbb{P}(Y_i = 1, T_i = 0 \mid Z_i = 0) - e(0)}{1 - \mathbb{P}(Y_i = 1, T_i = 0 \mid Z_i = 0) - e(0)} \right\} 
\leq \theta_{pr} \leq \min \left\{ \frac{\mathbb{P}(Y_i = 1, T_i = 1 \mid Z_i = 1) + 1 - e(1)}{1 - \mathbb{P}(Y_i = 1, T_i = 0 \mid Z_i = 0)}, 1 \right\}.
\]
ii. If \( P(Y_i = 0 \mid Z_i = 0) > 0 \) and \( P\{Y_i(1) = 1, \mathcal{E}\} \geq P\{Y_i(0) = 1, \mathcal{E}\} \) for \( \mathcal{E} \in \{NT, AT\} \), then the sharp identified interval of \( \theta_{pr} \) is given by

\[
\max\{0, \theta_L\} \leq \theta_{pr} \leq \min\left\{ \frac{P(Y = 1, T = 1 \mid Z = 1) + 1 - e(1)}{1 - P(Y = 1, T = 0 \mid Z = 0)}, 1 \right\}.
\]

Theorem 3 shows that \( \theta_L \) continues to be the sharp lower bound, provided that \( \theta_L \geq 0 \), that is, \( P(Y_i = 1 \mid Z_i = 1) \geq P(Y_i = 1 \mid Z_i = 0) \), and a stochastic dominance condition holds for the never-takers and always-takers. However, the upper bound will be larger than that in theorem 2 in general.

2. Identification with the Knowledge of the Exposure Rates

As in the case of DellaVigna and Kaplan (2007), the researcher may not directly observe \( T_i \) along with \( (Y_i, Z_i) \) but may have auxiliary data from which the exposure rates \( e(1) \) and \( e(0) \) can be estimated.\(^{10}\) In this case, the sharp identified bounds on \( \theta_{avg} \) become generally wider than those of theorem 2.

Assumption D (Observability of two marginals). Only the distribution of \( (Y_i, Z_i) \) and the exposure rates \( \{e(0), e(1)\} \) are known, where \( P(Y_i = 0 \mid Z_i = 0) > 0 \).

Theorem 4. Suppose that assumptions A, B, and D are satisfied. Then, the sharp identified interval of \( \theta_{pr} = \theta_{avg} \) is given by \([\theta_L, \theta_U]\), where \( \theta_L \) is given in equation (6) and

\[
\theta_U = \frac{\min\{1, P(Y_i = 1 \mid Z_i = 1) + 1 - e(1)\} - \max\{0, P(Y_i = 1 \mid Z_i = 0) - e(0)\}}{1 - \max\{0, P(Y_i = 1 \mid Z_i = 0) - e(0)\}}.
\]

Therefore, the upper bound in this case is nontrivial if and only if \( e(1) > P(Y_i = 1 \mid Z_i = 1) \).\(^{11}\) Note that it is the relative size of the take-up rate \( e(1) \) (e.g., the probability of reading a newspaper when a free subscription to it is offered) that determines how much we can hope to learn about the persuasion rate. For example, if the probability of watching Fox

\(^{10}\) The case in which the outcome and the treatment are separately observed belongs to an identification problem called the ecological inference problem. For instance, Cross and Manski (2002) and Manski (2018) discuss bounding a “long regression” by using information from a “short regression.” Their substantive concerns are distinct from ours.

\(^{11}\) The trivial case can occur in applications. See table 1 for such cases.
News is too small relative to the probability of voting for a Republican candidate when Fox News was introduced in the local cable, then it becomes difficult to pin down how successfully Fox News persuaded their audience to vote for a Republican candidate. Also, it is worth noting that $e(0) = 0$ is not uncommon, as table 1 and section VI show. In this case, the maximum in the expression of the upper bound is unnecessary. Intuition for the lower bound is the same as the case of theorem 2, because $\theta_L$ requires only the distribution of $Y_i$ given $Z_i$.

3. Identification with No Information

Associated with $T_i$

The final scenario is the least informative one, where $T_i$ is not observed at all. This is an almost trivial case, but we state it in a separate theorem for the sake of completeness.

**Assumption E (Limited observability).** No information associated with $T_i$ is available (i.e., the distribution of $(Y_i, Z_i)$ is all that is known), where $P(Y_i = 0 \mid Z_i = 0) > 0$.

**Theorem 5.** Suppose that assumptions A, B, and E are satisfied. Then, the sharp bound of $\theta_{pr} = \theta_{av}$ is given by $[\theta_L, \theta_U]$, where $\theta_L$ is given in equation (6).

The lower bound from theorem 4 depends only on the distribution of $(Y_i, Z_i)$, and therefore $\theta_L$ continues to be the lower bound in this case as well. Further, since no information for $e(1)$ and $e(0)$ is available, it suffices

**Table 1**

**Persuasion Rates: Papers on Voter Turnout**

| Paper                  | $\hat{y}(1)$ | $\hat{y}(0)$ | $\hat{e}(1)$ | $\hat{e}(0)$ | $\hat{f}$ | $[\theta_L, \theta_U]$ | $[\theta_{L*}, 1]$ |
|------------------------|--------------|--------------|--------------|--------------|-----------|-------------------------|-----------------|
| Green and Gerber 2000  | .472         | .448         | .279         | 0            | .156      | [0.043, 1]              | [0.086, 1]      |
| Green, Gerber, and     | .310         | .286         | .293         | 0            | .115      | [0.034, 1]              | [0.082, 1]      |
| Nickerson 2003         | .711         | .660         | .737         | 0            | .204      | [0.150, 0.924]          | [0.150, 1]      |
| Green and Gerber 2001  | .416         | .405         | .414         | 0            | .045      | [0.018, 1]              | [0.027, 1]      |
| Gentzkow 2006          | .455         | .435         | .800         | 0            | .044      | [0.035, 0.389]          | [0.035, 1]      |
| Gentzkow, Shapiro,     | .700         | .690         | .250         | 0            | .129      | [0.032, 1]              | [0.040, 1]      |
| and Sinkinson 2011     |              |              |              |              |           |                         |                 |

**Note.**—The outcome variable is voter turnout except for Gentzkow (2006), where exposure to television discouraged voters to go to the polls. Thus, for all rows to have positive persuasive effects, the outcome variable for Gentzkow (2006) is not to vote. In cols. 1 and 2, $\hat{y}(z)$ denotes estimates of $P(Y_i = 1 \mid Z_i = z)$ for $z = 1, 0$. $\hat{f}$ is the persuasion rate reported in DellaVigna and Gentzkow (2010, table 1). $[\theta_L, \theta_U]$ and $[\theta_{L*}, 1]$ are the sharp lower and upper bounds on the average and local persuasion rates, respectively, under assumptions A, B, and D. The third row corresponds to the row in table 1 of DellaVigna and Gentzkow (2010) under the treatment labeled “phone calls by youth.” The fourth row corresponds to the row in table 1 of DellaVigna and Gentzkow (2010) under the treatment labeled “phone calls 18–30-year-olds.” Table 1 of DellaVigna and Gentzkow (2010) gives $\hat{e}(1) - \hat{e}(0)$ but $\hat{e}(0) = 0$ in each row by study design.
to note that the upper bound in theorem 4 equals 1 whenever $e(0) > \mathbb{P}(Y_i = 1 \mid Z_i = 0)$ and $e(1) < \mathbb{P}(Y_i = 1 \mid Z_i = 1)$.

IV. The Local and Marginal Persuasion Rates

In this section, we consider rates of persuasion on other subpopulations that have been considered in econometrics: that is,

$$v_{\text{local}} := \mathbb{P}(Y_i(1) = 1 \mid Y_i(0) = 0, e(0) < V_i \leq e(1)),$$

$$v_{\text{marginal}}(v) := \mathbb{P}(Y_i(1) = 1 \mid Y_i(0) = 0, V_i = v) \text{ for } 0 < v < 1,$$

provided that the conditional probabilities are well defined: $v_{\text{local}}$ is the persuasion rate for the compliers (e.g., Imbens and Angrist 1994), whereas $v_{\text{marginal}}(v)$ is for the subpopulation such that $V_i = v$ (e.g., Heckman and Vytlacil 2005).

First, we obtain identification results for $v_{\text{local}}$ under the three sampling scenarios in the fuzzy persuasion design. The first step for this purpose is to note that the same reasoning as lemma 1 yields

$$v_{\text{local}} = \frac{\mathbb{P}(Y_i(1) = 1 \mid e(0) < V_i \leq e(1)) - \mathbb{P}(Y_i(0) = 1 \mid e(0) < V_i \leq e(1))}{1 - \mathbb{P}(Y_i(0) = 1 \mid e(0) < V_i \leq e(1))},$$

where the numerator is the LATE, which has received great attention in the econometrics literature (for a debate, see Deaton 2010; Heckman 2010; Imbens 2010). The denominator that rescales the LATE is also conditioned on the same subpopulation of the compliers.

Theorem 6. Suppose that assumptions A and B are satisfied.

i. Under assumption C, $v_{\text{local}}$ is point identified by $v_{\text{local}} = \theta^*$, where

$$\theta^* := \frac{\mathbb{P}(Y_i = 1 \mid Z_i = 1) - \mathbb{P}(Y_i = 1 \mid Z_i = 0)}{\mathbb{P}(Y_i = 0, T_i = 0 \mid Z_i = 0) - \mathbb{P}(Y_i = 0, T_i = 0 \mid Z_i = 1)}.$$

ii. Under assumption D, the sharp identified interval of $v_{\text{local}}$ is given by $[\theta^*_L, 1]$, where

$$\theta^*_L := \max \left\{ \theta_i, \frac{\mathbb{P}(Y_i = 1 \mid Z_i = 1) - \mathbb{P}(Y_i = 1 \mid Z_i = 0)}{e(1) - e(0)} \right\}.$$

iii. Under assumption E, the sharp identified interval of $v_{\text{local}}$ coincides with that of $\theta_{pr} = \theta_{avg}$, that is, $[\theta_L, 1]$.

Recall that the identification of the LATE requires the joint distribution of $(T_i, Z_i)$ and that of $(Y_i, Z_i)$ separately but not the full joint distribution of $(Y_i, T_i, Z_i)$. Unlike the LATE, the point identification in theorem 6(i)...
demands the knowledge of the joint distribution of \((Y_i, T_i, Z_i)\).\(^{12}\) Theorem 6(ii) shows that this requirement is not only sufficient but also necessary to achieve the point identification of \(\theta_{\text{local}}\).

Just like the LATE, it may be contentious whether \(\theta_{\text{local}}\) should be the parameter of interest, because the compliers are concerned with an unidentified subgroup of the population. However, we take a practical view that the identification results on \(\theta_{\text{local}}\) can complement the results obtained in section III.

The local persuasion rate \(\theta_{\text{local}}\) represents the average persuasive effect for a population that is different from the entire population. Given this caveat, it is interesting to note that in theorem 6(ii), the upper bound on \(\theta_{\text{local}}\) is always trivial in contrast to \(\theta_{\text{avg}}\), but the lower bound of \(\theta_{\text{local}}\) can never be worse than that of \(\theta_{\text{avg}}\). Therefore, in principle, the length of the identified interval of \(\theta_{\text{avg}}\) can be smaller than that of \(\theta_{\text{local}}\). If \(T_i\) is not observed at all, then there is no advantage in focusing on the compliers. Theorem 6(iii) confirms the intuition that the bounds for \(\theta_{\text{local}}\) are identical to those for \(\theta_{\text{avg}}\) if the distribution of \((Y_i, Z_i)\) is the only piece of information available. This corresponds to an uninteresting case for \(\theta_{\text{local}}\) though, as we have no information on compliers.

Data requirements for the identification of \(\theta_{\text{marginal}}(v)\) are generally quite demanding: for example, a continuous instrument is needed. However, identifying \(\theta_{\text{marginal}}(v)\) for various values of \(v\) can open up the possibility of point identification of \(\theta_{\text{avg}}\). Therefore, it is worth understanding what is sufficient for the identification of \(\theta_{\text{marginal}}\).

If \(Y_i\) and \(T_i\) are jointly observed along with a continuous instrument \(Z_i\), then \(\theta_{\text{marginal}}(v)\) can be point identified as in Heckman and Vytlacil (2005) and Carneiro, Heckman, and Vytlacil (2011). Examples of continuous instruments can be found in the literature on the media effects on voting. For instance, Enikolopov, Petrova, and Zhuravskaya (2011) and Della-Vigna et al. (2014) use the signal strength of NTV and Serbian radio as instruments, respectively; in both of the papers, \((Y_i, T_i, Z_i)\) are jointly observed. The following assumption describes the situation in which we can obtain point identification of \(\theta_{\text{marginal}}(v)\). We use the standard results in the literature (e.g., Heckman and Vytlacil 2005) for the subsequent theorem.

**Assumption F (Marginal treatment effects).**

i. The joint distribution of \((Y_i, T_i, Z_i)\) is known.

ii. \(T_i\) has the threshold structure in equation (5), where \(V_i\) is uniformly distributed on \([0, 1]\) and \(Z_i\) is independent of \((Y_i(t), V_i)\) for \(t = 0, 1\).

\(^{12}\) The denominator of eq. (11) requires that we know the marginal distribution of \(Y(0)\) for the compliers. Imbens and Rubin (1997) show that the marginal distributions of \(Y(1)\) and \(Y(0)\) for the compliers are identified if the joint distribution of \((Y, T, Z)\) is known; however, they did not consider the local persuasion rate.
iii. The distribution of $e(Z_i)$ is absolutely continuous with respect to Lebesgue measure.

**Theorem 7.** Suppose that assumptions A and F are satisfied. Then, for $v$ such that $v$ is in the interior of the support of $e(Z_i)$, $\theta_{\text{marginal}}(v)$ is point identified by

$$\theta_{\text{marginal}}(v) = \frac{\partial \mathbb{P}\{Y_i = 1 \mid e(Z_i) = e\}}{1 + |\partial \mathbb{P}\{Y_i = 1, T_i = 0 \mid e(Z_i) = e\}|},$$

provided that $\mathbb{P}\{Y_i = 1 \mid e(Z_i) = e\}$ and $\mathbb{P}\{Y_i = 1, T_i = 0 \mid e(Z_i) = e\}$ are continuously differentiable with respect to $e$.

Similar to the case of $\theta_{\text{avg}}$ or $\theta_{\text{local}}$, assumption A enables us to rewrite $\theta_{\text{marginal}}(v)$ as $E\{Y_i(1) - Y_i(0) \mid V_i = v\}/\mathbb{P}\{Y_i(0) = 0 \mid V_i = v\}$; that is, $\theta_{\text{marginal}}(v)$ is a rescaled version of the marginal treatment effect of Heckman and Vytlacil (2005). Theorem 7 is a direct consequence of that.

Theorem 7 does not consider the other two scenarios of data availability. This is mainly because continuous instruments are relatively infrequent in the context of persuasion, and we are not aware of any applications where continuous instruments are available while the outcome and treatment are not jointly observed.

If the support of the exposure rate $e(Z_i)$ is equal to the unit interval $[0, 1]$, then theorem 7 shows the identification of $\theta_{\text{marginal}}(v)$ for all $v$ in the unit interval. Then, we can use $\theta_{\text{marginal}}(v)$ to construct different policy-oriented quantities, as in Heckman and Vytlacil (2005) and Carneiro, Heckman, and Vytlacil (2011). For instance, the persuasion rate of the entire population can be obtained by $\int_0^1 \theta_{\text{marginal}}(v) dF\{v \mid Y_i(0) = 0\}$, which is equal to

$$\theta_{\text{avg}} = \frac{\int_0^1 \partial \mathbb{P}\{Y_i = 1 \mid e(Z_i) = e\}}{1 + \int_0^1 |\partial \mathbb{P}\{Y_i = 1, T_i = 0 \mid e(Z_i) = e\}|} dv$$

by Bayes’s theorem.

**V. Discussion**

In this section, we articulate the relationships between $\theta_{\text{avg}}$, $\theta_{\text{local}}$, and the DK measures $f$ and $\tilde{f}$ defined in section II, and we summarize the main takeaways of our identification results. We assume that assumptions A and B hold throughout this section, so we have $\theta_{\text{pr}} = \theta_{\text{avg}}$. Also, in order to be consistent with the identification analysis, we work with the population versions of $f$ and $\tilde{f}$: that is,
persuasion does estimate the persuasion rate unless any of them are ill defined, if no one is affected by the information, that for the compliers, which is clear from the fact that

\[
P\{Y_i(0) = 0\} \theta_{DK} = P\{Y_i = 0 | Z_i = 0\} \tilde{\theta}_{DK}
\]

\[
= P\{Y_i(0) = 0 | e(0) < V_i \leq e(1)\} \theta_{local}
\]
is equal to the LATE, while \(P\{Y_i(0) = 0\} \theta_{avg}\) is equal to the ATE. This is the same as the ATE when any of the cases applies. As a result, we have that in case i, \(\theta_{DK} = \theta_{DK} = \theta_{avg} = \theta_{local}\) holds; in case ii, \(\theta_{DK} = \theta_{avg} = \theta_{local} \leq \theta_{DK}\); and in case iii, \(\theta_{DK} = \theta_{avg} = \theta_{local} = 1 \leq \theta_{DK}\) if everybody is persuaded, or \(\theta_{DK} = \theta_{DK} = \theta_{avg} = \theta_{local} = 0\), unless any of them are ill defined, if no one is affected by the information. Therefore, the feasible version \(\tilde{f}\) of the DK measure of persuasion does estimate the persuasion rate \(\theta_{avg}\) in case i, although as DellaVigna and Kaplan (2007) correctly pointed out, \(\theta_{DK}\) does approximate \(\theta_{DK}\) in case ii as well if either \(e(0)\) or \(\theta\) is close to zero.

Our identification results in section III show that \(\theta_{L}\) is the sharp lower bound of the identified interval of \(\theta_{avg}\) in the fuzzy design regardless of whether the full joint distribution of the outcome, treatment, and instrument

\[\text{f} \stackrel{p}{\longrightarrow} \theta_{DK} \equiv \frac{P(Y_i = 1 | Z_i = 1) - P(Y_i = 1 | Z_i = 0)}{e(1) - e(0)} \frac{1}{1 - P\{Y_i(0) = 1\}}. \tag{13}\]

\[\tilde{f} \stackrel{p}{\longrightarrow} \tilde{\theta}_{DK} \equiv \frac{P(Y_i = 1 | Z_i = 1) - P(Y_i = 1 | Z_i = 0)}{e(1) - e(0)} \frac{1}{1 - P\{Y_i(1) = 1 | Z_i = 0\}}. \tag{14}\]

First, neither \(\theta_{DK}\) nor \(\tilde{\theta}_{DK}\) is generally equal to the persuasion rate \(\theta_{avg}\) or that for the compliers, \(\theta_{local}\), which is equal to the LATE. For example, \(\theta_{DK}\) rescales the LATE with an unconditional probability and hence it does not render a well-defined conditional probability in general. Similarly, \(\tilde{\theta}_{DK}\) is not necessarily a rate in spite of the rescaling factor. So, making comparisons across different studies on the basis of \(\theta_{DK}\) or \(\tilde{\theta}_{DK}\) can be misleading, although it is a common practice (e.g., DellaVigna and Gentzkow 2010).

There are some special cases of exception though: (i) everybody is a complier, as in the sharp persuasion design; (ii) \(T_i\) is independent of the potential outcomes \(Y_i(\cdot)\) for \(t = 0, 1\), conditional on \(Z_i\); or (iii) there is no heterogeneity in the treatment effect in that \(Y_i(1) - Y_i(0)\) is a constant. That is, in cases i and ii, there is no endogeneity issue, whereas in case iii, no one is affected by the persuasive message (i.e., \(Y_i(1) - Y_i(0) = 0\) for all \(i\)), or everybody is persuaded (i.e., \(Y_i(1) - Y_i(0) = 1\) for all \(i\)). Note that the LATE is the same as the ATE when any of the three cases applies. As a result, we have that in case i, \(\theta_{DK} = \tilde{\theta}_{DK} = \theta_{avg} = \theta_{local}\) holds; in case ii, \(\theta_{DK} = \theta_{avg} = \theta_{local} \leq \theta_{DK}\); and in case iii, \(\theta_{DK} = \theta_{avg} = \theta_{local} = 1 \leq \theta_{DK}\) if everybody is persuaded, or \(\theta_{DK} = \theta_{DK} = \theta_{avg} = \theta_{local} = 0\), unless any of them are ill defined, if no one is affected by the informational treatment. Therefore, the feasible version \(\tilde{f}\) of the DK measure of persuasion does estimate the persuasion rate \(\theta_{avg}\) in case i, although as DellaVigna and Kaplan (2007) correctly pointed out, \(\theta_{DK}\) does approximate \(\theta_{DK}\) in case ii as well if either \(e(0)\) or \(\theta\) is close to zero.

Our identification results in section III show that \(\theta_{L}\) is the sharp lower bound of the identified interval of \(\theta_{avg}\) in the fuzzy design regardless of whether the full joint distribution of the outcome, treatment, and instrument

\[13\) Recall that the population version of the Wald statistic is equal to the LATE under assumption B.

\[14\) In case ii, we have \(P(Y_i = 1 | Z_i = 0) = P(Y_i = 1, T_i = 1 | Z_i = 0) + P(Y_i = 1, T_i = 0 | Z_i = 0) = P\{Y_i(0) = 1\} + [P\{Y_i(1) = 1\} - P\{Y_i(0) = 1\}]e(0)\).
is available. The parameter $\theta_L$ has been reported in the literature without understanding that it is the sharp lower bound on $\theta_{avg}$. For instance, DellaVigna and Gentzkow (2010) extensively estimate $\theta_{DK}$ by using many examples, but they report $\theta_L$ as a lower bound of $\hat{\theta}_{DK}$ when $T_i$’s are unobserved and hence $e(1)$ and $e(0)$ are unknown. Our results show that $\theta_L$ is always a meaningful parameter, but $\hat{\theta}_{DK}$ may not. Therefore, even when information about $e(0)$ and $e(1)$ is available, $\theta_L$ is a better parameter to estimate than $\hat{\theta}_{DK}$.

Indeed, if the full joint distribution of $(Y_i, T_i, Z_i)$ is available, then we recommend reporting $[\theta_L, \theta_U]$ along with $\theta^*$; these can be consistently estimated by their sample analogs. If $(Y_i, Z_i)$ is observed with some auxiliary information for $e(0)$ and $e(1)$ available, then $[\theta_L, \theta_U]$ and $[\theta^*_L, 1]$ should be reported. If $T_i$ is not observed at all, then the interval $[\theta_L, 1]$ is the best we can hope for to study either $\theta_{avg}$ or $\theta_{local}$.

Note that $\theta_L$ should be estimated all the time; it requires data only on $(Y_i, Z_i)$. Because the actual $T_i$ can be difficult to observe, researchers have used an extra micro-level survey to obtain auxiliary data on $T_i$ which seems quite costly. However, the value of an attempt to observe $T_i$ can be limited, depending on which parameter the researcher wants to learn about. For instance, if the researcher cares about the persuasion rate of the entire population, then observing $T_i$ does not add any information for the lower bound, while it can potentially improve the upper bound. If the group of compliers is of interest, then whether we observe $T_i$ and how we observe it can be relevant issues; we have $\theta^*_L \geq \theta_L$ in the second data scenario, and $\theta^*$ is point identified if $(Y_i, T_i, Z_i)$ is jointly observed. If $Z_i$ is continuously distributed, the value of observing $(Y_i, T_i, Z_i)$ jointly increases dramatically as well. In summary, our identification analysis shows that the value of observing $T_i$ depends crucially on which population the researcher is interested in.

In order to illustrate the difference between the DK measure and our bounds, we have calculated them in table 1. We focus on the results reported in DellaVigna and Gentzkow (2010, their table 1) when the outcome variable is voter turnout. We have chosen this type of study because the turnout is among the most studied outcome variables in the literature and it is naturally a binary measure. Table 1 provides estimates of $P(Y_i = 1 \mid Z_i = z)$ and $e(z)$ for $z = 0, 1$, thereby enabling us to obtain the bounds based on theorem 4 and theorem 6(ii). It can be seen that using DellaVigna and Kaplan’s (2007) persuasion rates alone may lead to misleading conclusions because the bounds on $\theta_{avg}$ as well as those on $\theta_{local}$ are in fact wide. Moreover, the results in table 1 suggest that identification power under assumptions A, B, and D in this example is limited, especially for the upper bounds on $\theta_{avg}$ and $\theta_{local}$. We further illustrate these points with empirical examples in section VI.

Finally, since the parameters are partially identified, inference should also account for that. The method proposed by Stoye (2009) is useful for
that purpose, at least in the most favorable data scenario, in which case the sample analog principle and the delta method show that we can construct the estimators \( \hat{\theta}_L \) and \( \hat{\theta}_U \) that are asymptotically jointly normal. Therefore, by Stoye (2009), a \((1 - \alpha)\) confidence interval for \( \theta_{avg} \) can be obtained by \([\hat{\theta}_L - c_w \hat{\sigma}_L, \hat{\theta}_U + c_w \hat{\sigma}_U]\), where \( \hat{\sigma}_L \) and \( \hat{\sigma}_U \) are the estimated standard errors of \( \hat{\theta}_L \) and \( \hat{\theta}_U \), respectively, and \( c_w \) is chosen by solving

\[
\Phi\left(c_w + \frac{\hat{\Delta}}{\max(\hat{\sigma}_L, \hat{\sigma}_U)}\right) - \Phi(-c_w) = 1 - \alpha,
\]

where \( \Phi \) is the distribution function of the standard normal and \( \hat{\Delta} \) is the estimated length of the identified interval.

The second data scenario is slightly more complicated, because \( \theta_U \) and \( \theta_L^* \) contain the min or max function that is not smooth, so the delta method does not apply. In appendixes F and G, we propose a two-step method for inference to overcome this problem, which we have applied to the empirical example we discuss in section VI. In the third data scheme, confidence intervals for \( \theta_{avg} \) and \( \theta_{local} \) always coincide, and they can be obtained by using a one-side critical value on \( \hat{\theta}_L \). Specifically, they are given by \([\hat{\theta}_L - z_{1-\alpha} \hat{\sigma}_L, 1]\), where \( z_{1-\alpha} \) is the \((1 - \alpha)\) quantile of the standard normal distribution. Appendixes F and G provide a more detailed discussion on inference. Furthermore, see appendix H for semiparametrically efficient estimation of the two key parameters, that is, \( \theta_L^* \) and \( \theta_L^* \), when exogenous covariates \( X_i \) are present and integrated out.

In sum, this paper clarifies identification issues when we insert exposure as a choice variable and employ a proper causal framework that is used in policy evaluation to model two causal links (i.e., \( Z_i \rightarrow T_i \) and \( T_i \rightarrow Y_i \)).

VI. Empirical Examples

A. Effects of Uncensored Media

In this section, we revisit Chen and Yang (2019), who conducted a field experiment in China to measure the effects of providing students with internet access to the uncensored media on various outcome variables. Excluding the existing users, the subjects in their experiments consist of four groups: (i) the control group, (ii) the control group of students who were encouraged to visit foreign news websites blocked by the Great Firewall, (iii) students who received free access to uncensored internet, and (iv) students who received both the access and encouragement treatments. They followed the subjects over 18 months to collect outcomes on media-related behaviors, beliefs, and attitudes among other things.

\[\text{15 We are grateful to an anonymous referee who provided us with insightful comments.}\]
It turns out that there were no differences between groups i and ii and the effects were the largest for group iv, that is, the access plus encouragement group (hereafter, Group-AE students). To benchmark their findings, Chen and Yang (2019) computed the DK measure of persuasion for the Group-AE students (see table A.13 of their paper for details) and commented that “their estimated persuasion rates are of a similar magnitude to those found in authoritarian regimes that typically have highly regulated media markets” (Chen and Yang 2019, 2323–24).

In this section, we use the data from Chen and Yang (2019) to illustrate how the common practice of reporting the DK measure can lead to misleading conclusions by contrasting the DK measure of persuasion with our proposed approaches. As in Chen and Yang (2019), we focus on the Group-AE students. That is, $Z_i = 1$ if the $i$th subject is randomly assigned to the Group-AE group, and $Z_i = 0$ if the $i$th subject is randomly assigned to the control or control-encouragement group while dropping the access-only group and the existing users. In Chen and Yang’s (2019) two-stage least squares analysis (table 3 in their paper), the treatment variable is $T_i = 1$ if the $i$th subject is an active user of the censorship circumvention tool and $T_i = 0$ otherwise. We use the same treatment variable in our analysis. To replicate the results in Chen and Yang (2019) in a representative but succinct way, we focus on the 11 outcome variables listed in panel A of table A.13 in their paper. They represent media-related behaviors, beliefs, and attitudes and are transformed to binary variables by Chen and Yang (2019).

Recall the population version of the DK measure $f$ (see eq. [13]). In their table A.13, Chen and Yang (2019) measure $\mathbb{P}(Y_i = 1 \mid Z_i = 1) - \mathbb{P}(Y_i = 1 \mid Z_i = 0)$ by the ITT effects of the Group-AE assignment and approximate $\mathbb{P}\{Y_i(0) = 1\}$ using variables collected at the time of the baseline survey or by the estimates of $\mathbb{P}(Y_i = 1 \mid Z_i = 0)$ as in $\hat{\theta}_{0|k}$, if the former is unavailable. Table 2 in Chen and Yang (2019) and the data provided by the authors indicate that the change in the exposure rate is 45.5% if the treatment status is measured by being active users, and therefore we use $e(1) - e(0) = 0.455$ for our subsequent calculations.

Table 2 summarizes the empirical results. In column 1 of table 2, we recompute Chen and Yang’s (2019) persuasion rates for the 11 outcome variables listed in panel A of table A.13 in their paper. As we explained above, these estimates are based on $e(1) - e(0) = 0.455$. Out of the 11 outcome variables, the median persuasion rate is 101%, and therefore these persuasion rates cannot be understood as conditional probabilities. Columns 2 and 3 report our estimates of the average and local persuasion rates along with the 95% confidence intervals (in curly braces) that were obtained via 10,000 bootstrap replications. Our estimates show that (i) the average persuasion rates are only partially identified, and the widths of the identified intervals are substantial for most of
(ii) the point-identified local persuasion rates are contained by the identified intervals for the average persuasion rates and are typically closer to the upper end points of the intervals. The persuasive effects are of a relatively large magnitude in that the smallest lower end point of the confidence interval for the average persuasion is 16%. However, the original estimates overstate the magnitude by a substantial factor and mask underidentification of average persuasion rates. In short, we find that the subjects in the experiments responded to exposure to uncensored internet highly heterogeneously, indicating that it is important to go beyond the benchmark measures of the DK type.

| OUTCOME | CHEN AND YANG 2019 (1) | PERSUASION RATES |  |
|---------|-----------------------|------------------|---|
|         |                       | Average (2)      | Local (3) |
| A.1.2. Ranked high: foreign websites | 35.5 | [21.9, 68.4] | 41.0 |
|         |                       | [15.6–72.9]      | [29.2–51.1] |
| A.1.6. Frequency of visiting foreign websites for information | 168.5 | [55.5, 76.6] | 70.4 |
|         |                       | [50.6–80.2]      | [64.6–75.7] |
| A.2.1. Purchase discounted tool we offered | 45.6 | [22.7, 64.2] | 38.7 |
|         |                       | [19.3–67.9]      | [32.6–44.9] |
| A.2.2. Purchase any tool | 101.2 | [49.5, 85.0] | 76.8 |
|         |                       | [45.5–87.7]      | [71.9–81.6] |
| A.3. Valuation of access to foreign media outlets | 121.7 | [49.3, 82.1] | 73.4 |
|         |                       | [42.8–85.8]      | [65.8–80.0] |
| A.4. Trust in nondomestic media outlets | 158.1 | [62.9, 85.8] | 81.6 |
|         |                       | [57.8–89.0]      | [76.2–86.4] |
| A.5.1. Degree of censorship on domestic news outlets | 82.3 | [35.1, 73.5] | 57.0 |
|         |                       | [29.6–77.4]      | [48.7–64.3] |
| A.5.2. Degree of censorship on foreign news outlets | 114.6 | [39.1, 74.4] | 60.4 |
|         |                       | [33.8–78.2]      | [53.1–67.3] |
| A.6. Censorship unjustified | 77.6 | [31.6, 66.6] | 48.6 |
|         |                       | [23.5–72.0]      | [35.8–59.2] |
| A.7.1. Domestic censorship driven by government policies | 192.7 | [75.0, 88.6] | 86.8 |
|         |                       | [64.4–94.2]      | [76.5–94.4] |
| A.7.2. Foreign censorship driven by government policies | 54.1 | [36.9, 76.3] | 61.0 |
|         |                       | [19.2–84.1]      | [32.4–77.4] |

Note.—In col. 1, we recompute Chen and Yang’s (2019) persuasion rates for the 11 outcome variables listed in panel A of table A.13 in their paper. For these estimates, we use $e(1) - e(0) = 0.455$. Column 2 reports the bound estimates of the average persuasion rates. Column 3 reports the point estimates of the local persuasion rates. The 95% confidence intervals are given in curly braces, obtained via 10,000 bootstrap replications. The persuasion rates are expressed in percentages.
B. Effects of Political News

We now illustrate our proposed methods by using data from Gerber, Karlan, and Bergan (2009), who report findings from a field experiment to measure the effect of political news. We have chosen this example because it contains a credible binary instrument from the field experiment and we can also illustrate all of the three sampling scenarios as well as the case of nonbinary outcomes; for the theory on the multinomial outcome case, see appendix B. In Gerber, Karlan, and Bergan (2009), there are three statuses in the ITT: a control group, an offer of a free subscription to the Washington Post, and an offer of a free subscription to the Washington Times. To illustrate the usefulness of our paper, we focus on the Washington Post and drop all observations from the Washington Times subscription. That is, \( Z_i = 1 \) if the \( i \)th individual received a free subscription to the Washington Post and \( Z_i = 0 \) if not.

Focusing on the ITT analysis, Gerber, Karlan, and Bergan (2009) have reported ITT estimates for various outcomes \( Y_i \). DellaVigna and Gentzkow (2010) compute persuasion rates for Gerber, Karlan, and Bergan (2009), for which they simply set \( T_i = 1 \) if the \( i \)th individual opted in to the free subscription and \( T_i = 0 \) if they opted out of it.\(^{16}\) In this section, for the purpose of illustrating our identification results, we consider a different treatment variable: \( T_i = 1 \) if the \( i \)th individual read a newspaper at least several times per week and \( T_i = 0 \) otherwise, which is a variable that Gerber, Karlan, and Bergan (2009) kept track of in a follow-up survey. Therefore, the relevant treatment we consider differs from that of Della-Vigna and Gentzkow (2010): it is whether individuals have actually read the newspaper. The outcome variables we consider are as follows. For the binary case, \( Y_i = 1 \) if the \( i \)th individual reported voting for the Democratic candidate in the 2005 gubernatorial election, and \( Y_i = 0 \) if the subject did not vote for the Democratic candidate or did not vote at all. For the multinomial case, not voting at all is treated as an outside option. We use only a subsample of the data from Gerber, Karlan, and Bergan (2009) with those who responded to the follow-up survey to use information on \((Y_i, T_i, Z_i)\) jointly. After dropping observations for the Washington Times subscription and removing missing data, we summarize the data from Gerber, Karlan, and Bergan (2009) in table 3. Although the joint distribution of \((Y_i, T_i, Z_i)\) is observed in this example, we also consider using the two marginals of \((Y_i, Z_i)\) and \((T_i, Z_i)\) separately to make a comparison. The estimates are summarized in table 4. Because the size of the sample extract we use is relatively modest (\( n = 701 \)) for an interval-identified

---

\(^{16}\) In app. C, we provide the empirical results of bound analysis using the treatment variable for opting into the free subscription.
object, we report the 80% confidence intervals obtained by the inference methods described in section V as well as in appendixes F and G.

First, we discuss the case where the full joint distribution of \((Y_i, T_i, Z_i)\) is used. In this data scenario, the average effect of persuasion by reading the newspaper is bounded between 7% and 63%. In contrast, the persuasion rate for the group of compliers is point estimated by 81%. It is interesting to note that the estimate of \(v_{local}\) is so large that it is greater than the upper bound of \(v_{avg}\). This suggests that individuals are highly heterogeneous in this example, indicating that \(\theta_{DK}\) might not be a well-defined conditional probability here. Indeed, the estimate of \(\theta_{DK}\) in equation (14) is \(\theta_{DK} = 1.1027\), which is greater than 1.

When the marginals of \((Y_i, Z_i)\) and \((T_i, Z_i)\) are used separately, the upper bound on \(v_{avg}\) increases from 63% to 78%. Further, \(v_{local}\) is no longer point estimated, but we only know that it is bounded between 78% and 81%.

### TABLE 3
**Summary Statistics of Data from Gerber, Karlan, and Bergan (2009)**

| Voted for Democrat | \(T_i = 0\) | \(T_i = 1\) | Total |
|--------------------|-------------|-------------|-------|
| **Washington Post (Z_i = 1):** | | | |
| \(Y_i = 0\)       | 94          | 93          | 187   |
| \(Y_i = 1\)       | 31          | 68          | 99    |
| **Total**         | 125         | 161         | 286   |
| **Control (Z_i = 0):** | | | |
| \(Y_i = 0\)       | 162         | 150         | 292   |
| \(Y_i = 1\)       | 46          | 77          | 123   |
| **Total**         | 208         | 207         | 415   |

**Note.**—Data from Gerber, Karlan, and Bergan (2009) are used after dropping observations for the *Washington Times* subscription and removing missing data.

### TABLE 4
**Estimates of Key Parameters**

| \((Y_i, T_i, Z_i)\) | \((Y_i, Z_i)\) and \((T_i, Z_i)\) | \((Y_i, Z_i)\) Only |
|---------------------|----------------------------------|---------------------|
| **ITT**             | 0.498                            | [0.0036, 0.0959]    |
| \(\theta_{avg}\)   | [.0707, .6343]                   | [.0707, .7832]      |
|                     | [0.0289, 0.6610]                 | [.0286, 0.8143]     |
| **LATE**            | .7759                            | [.0498, 1]          |
|                     | [0–1]                            | [0.0195–1]          |
| \(\theta_{local}\) | .8067                            | [.7759, 1]          |
|                     | [0.1243–1]                       | [0.069–1]          |

**Note.**—The first and third rows show the estimates of ITT and LATE, respectively. The second row corresponds to \([\hat{\theta}_1, \hat{\theta}_1]\), \([\hat{\theta}_1, \hat{\theta}_1]\), and \([\hat{\theta}_1, 1]\), respectively. The fourth row shows \(\hat{\theta}_L, [\hat{\theta}_L, 1]\), and \([\hat{\theta}_L, 1]\), respectively. The 80% confidence intervals are given in curly braces. The sharp identified interval for LATE is \([ITT, 1]\) when only the joint distribution of \((Y_i, Z_i)\) is available. This is because \(\epsilon(1) – \epsilon(0)\) is unknown and can be everywhere between 0 and 1 in this case.
100%. This difference illustrates the loss of identification power if we do not observe the joint distribution of \((Y_i, T_i, Z_i)\).

Finally, we estimate the lower bound on the average persuasion rate by additionally conditioning on those who would vote even without reading the newspaper (see app. B for details). The resulting lower bound on the average persuasion increases from 0.0707 (0.0289) to 0.0975 (0.0554), where the numbers in parentheses are the left-end points of the 80% confidence intervals. Therefore, the (point-identified) ITT effect is 5%, while the lower bound of the average persuasion rate is about 7%, or 10% if we further focus on those who would vote without reading the newspaper.

**VII. Conclusions**

We have set up a simple econometric model of persuasion, introduced several parameters of interest, and analyzed their identification. The empirical examples in sections V and VI as well as the examples in appendixes D and E demonstrate that the persuasive effects are highly heterogeneous in the settings of media and fundraising.

We have focused on the case of binary outcomes and binary treatments. In appendix B, we extend our analysis to nonbinary outcomes. If the outcome is nonbinary, then we can condition on those who would not choose the outside option without the treatment. For instance, suppose that we have three options of voting for a Republican, voting for a Democrat, or not voting at all. Then, the persuasive effect of a message supporting a Republican can be measured in a couple of different ways: focusing on those who would not have voted for a Republican without the message is one way, and conditioning on those who would have voted for a Democrat (i.e., voted but not voted for a Republican) is the other. In the latter case, we show that the resulting lower bound is always no smaller than that of the binary outcome case.

In general, treatments are multivalued: unordered treatments (e.g., watching Fox News, CNN, or MSNBC) and ordered treatments (e.g., numbers of hours watching Fox News) arise naturally in applications. It would be fruitful to build on recent developments in multivalued treatments (e.g., Heckman, Urzua, and Vytlacil 2006, 2008; Heckman and Vytlacil 2007b; Heckman and Pinto 2018; Lee and Salanić, 2018) to investigate identification of persuasive effects. It would also be interesting to estimate deep parameters in an economic model of persuasion by using a more structural approach in the setup of multivalued treatments. These are topics for future research.

**Data Availability**

Code replicating the tables and figures in this article, including those presented in the appendixes, can be found in Jun and Lee (2023) in the Harvard Dataverse, https://doi.org/10.7910/DVN/YUHGD5.
References

Bassi, V., and I. Rasul. 2017. “Persuasion: A Case Study of Papal Influence on Fertility Preferences and Behavior.” American Econ. J. Appl. Econ. 9 (4): 250–302.

Bergemann, D., and S. Morris. 2019. “Information Design: A Unified Perspective.” J. Econ. Literature 57 (1): 44–95.

Calvi, R., A. Lewbel, and D. Tommasi. 2022. “LATE with Missing or Mismeasured Treatment.” J. Business and Econ. Statis. 40 (4): 1701–17.

Carneiro, P., J. J. Heckman, and E. Vytlacil. 2011. “Estimating Marginal Returns to Education.” A.E.R. 101 (6): 2754–81.

Che, Y.-K., W. Dessein, and N. Kartik. 2013. “Pandering to Persuade.” A.E.R. 105 (1): 47–79.

Chen, Y., and D. Y. Yang. 2019. “The Impact of Media Censorship: 1984 or Brave New World?” A.E.R. 109 (6): 2294–332.

Cross, P. J., and C. F. Manski. 2002. “Regressions, Short and Long.” Econometrica 70 (1): 357–68.

Deaton, A. 2010. “Instruments, Randomization, and Learning about Development.” J. Econ. Literature 48 (2): 424–55.

DellaVigna, S., R. Enikolopov, V. Mironova, M. Petrova, and E. Zhuravskaya. 2014. “Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia.” American Econ. J. Appl. Econ. 6 (3): 103–32.

DellaVigna, S., and M. Gentzkow. 2010. “Persuasion: Empirical Evidence.” Ann. Rev. Econ. 2:643–69.

———. 2007. “The Fox News Effect: Media Bias and Voting.” Q.J.E. 122 (3): 1187–234.

Enikolopov, R., M. Petrova, and E. Zhuravskaya. 2011. “Media and Political Persuasion: Evidence from Russia.” A.E.R. 101 (7): 3253–85.

Gentzkow, M. 2006. “Television and Voter Turnout.” Q.J.E. 121 (3): 931–72.

Gentzkow, M., and E. Kamenica. 2017. “Competition in Persuasion.” Rev. Econ. Studies 84 (1): 300–322.

Gentzkow, M., J. M. Shapiro, and M. Sinkinson. 2011. “The Effect of Newspaper Entry and Exit on Electoral Politics.” A.E.R. 101 (7): 2980–3018.

Gerber, A. S., D. Karlan, and D. Bergan. 2009. “Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions.” American Econ. J. Appl. Econ. 1 (2): 35–52.

Green, D. P., and A. S. Gerber. 2000. “The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment.” American Polit. Sci. Rev. 94 (3): 653–63.

———. 2001. “Getting Out the Youth Vote: Results from Randomized Field Experiments.” Natural Field Experiments no. 00260, Field Experiments.

Green, D. P., A. S. Gerber, and D. W. Nickerson. 2003. “Getting Out the Vote in Local Elections: Results from Six Door-to-Door Canvassing Experiments.” J. Politics 65 (4): 1083–96.

Heckman, J. J. 2010. “Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy.” J. Econ. Literature 48 (2): 356–98.

Heckman, J. J., and R. Pinto. 2018. “Unordered Monotonicity.” Econometrica 86 (1): 1–35.

Heckman, J. J., J. Smith, and N. Clements. 1997. “Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts.” Rev. Econ. Studies 64 (4): 487–555.

Heckman, J. J., S. Urzua, and E. Vytlacil. 2006. “Understanding Instrumental Variables in Models with Essential Heterogeneity.” Rev. Econ. and Statis. 88 (3): 389–432.
