Planning the Optimal Get-out-the-vote Campaign Using Randomized Field Experiments*

Kosuke Imai† Aaron Strauss‡

First Draft: December 4, 2008
This Draft: May 11, 2009

Abstract

Political scientists have recently conducted over a hundred of randomized field experiments to examine the effectiveness of various mobilization methods for increasing voter turnout. Given the high degree of internal and external validity, the empirical findings of these studies have a potential to significantly impact the practice of get-out-the-vote (GOTV) campaigns in the real world. In this paper, we offer an essential and yet missing methodological tool that allows GOTV campaign planners to best utilize the results of such field experiments. In particular, we show how to derive the optimal GOTV campaign strategy from field experiments. Our nonparametric method is applicable to partisan or nonpartisan campaigns as well as campaigns with multiple mobilization methods of the same or different costs. We evaluate the effectiveness of the proposed method using three existing field experiments. In multiple cases, we find that the resulting optimal campaign strategy is more than twice as cost-effective as a naive strategy.

*Financial support from the National Science Foundation (SES–0550873 and SES–0752050) and the Princeton University Committee on Research in the Humanities and Social Sciences is acknowledged. We thank useful comments from seminar participants at Columbia University and the University of Wisconsin, Madison.

†Assistant Professor, Department of Politics, Princeton University, Princeton NJ 08544. Phone: 609–258–6610, Fax: 973–556–1929, Email: kimai@princeton.edu, URL: http://imai.princeton.edu

‡Ph.D. student, Department of Politics, Princeton University, Princeton NJ 08544

1 Introduction

Over the last decade, political scientists have shown renewed interests in the use of randomized field experiments to study voter turnout (e.g., Gerber and Green, 2000; Nickerson, 2007, among many others). Building on the previous work from more than a half century ago (e.g., Gosnell, 1927; Hartmann, 1936; Eldersveld, 1956), these researchers have developed creative ways to conduct field experiments with a large number of voters in real electoral environment in order to test various theories of voter turnout (e.g., Gerber *et al.*, 2003; Nickerson, 2008). Given the high degree of internal and external validity, the empirical findings of these studies have a potential to significantly impact the practice of get-out-the-vote (GOTV) campaigns in the real world (Green and Gerber, 2008).

Despite the accumulating empirical evidence, political science articles reporting the results of field experiments have a strong tendency to focus on the statistical significance of the estimated *overall* average treatment effect (ATE) of each mobilization technique. In addition, many researchers implicitly assume the constant additive treatment effect across individual voters when using regression models for statistical analysis. This leads to the standard practice of reporting a single estimate (or at most a small number of estimates) summarizing the efficacy of each mobilization method.

However, the findings based on such an approach may not be of much use for GOTV campaign planners for two reasons. First, a planner must consider the problem of *treatment effect heterogeneity* where each of available GOTV techniques may mobilize different voters to a varying degree. The second and related problem is that a planner faces a budget constraint and must evaluate the *cost-effectiveness* of available mobilization strategies which itself may vary across different voters. Thus, uniformly applying a single mobilization method to the entire target population is at best suboptimal and sometimes is not even feasible. This means that the standard practice of only reporting the estimated overall ATE by academic researchers may not provide useful information from a practitioner's perspective. Similarly, the common assumption of constant additive treatment effect is too restrictive and does not serve the purpose of campaign planners.

In this paper, we address this gap between academic research and policy making by offering an essential and yet missing methodological tool. In particular, we propose a formal decision-theoretic framework that allows GOTV campaign planners, both partisan and nonpartisan, (or academic researchers who wish to inform the planners through their experimental research) to best utilize the results of field experiments when planning their own mobilization campaigns. In the proposed Bayesian decision theoretic framework, a nonpartisan planner maximizes the posterior expected turnout among a target population of voters subject to a budget constraint, whereas the objective function of a partisan planner is the posterior probability that the party's own candidate wins the election. A planner starts with a prior belief about the effectiveness of each mobilization technique under consideration for voters with different characteristics and updates this belief based on the available experimental data. Under this setting, we show how to derive the optimal GOTV campaign strategy for the target population of voters using the data from randomized field experiments.

The problem of optimizing campaign tactics across voter characteristics was first introduced by Kramer (1966) more than forty years ago. In that article, a campaign manager is assumed to face the decision of whether to conduct blind or selective canvassing across precincts subject to a budget constraint. After illustrating his decision-theoretic approach with artificial data about the partisan make-up of precincts, Kramer (1966) concludes by arguing,

> the use of quantitative methods for policy analysis has proved to be fruitful in many different fields, and these methods deserve to be more widely known, and used, in

political science (p.160).

Unfortunately, few scholars who conduct GOTV field experiments have followed up on Kramer's proposal to inform policy makers with a formal decision-theoretic analysis. One exception is a small number of articles that have examined a related question of which subgroups exhibit larger treatment effects (e.g., Gerber, 2004; Nickerson and Arceneaux, 2006; Arceneaux and Kolodny, 2007). These findings may be used by campaign managers as a basis for planning their GOTV campaign.

However, an important and well-known methodological problem is that if subgroups are formed after the experiment is conducted in an ad hoc manner, such an analysis runs a risk of finding statistically significant results when no true relationship exists (e.g., Pocock *et al.*, 2002). Moreover, these studies do not examine the issue of cost effectiveness in the presence of a budget constraint, which is an essential consideration for campaign planners. Thus, we need a principled and systematic approach that can be used to derive the optimal campaign strategy from experimental data while avoiding the post-hoc subgroup analysis and other relevant statistical problems.

Our proposed methodology extends Kramer's pioneering work by placing the campaign planner's problem in the formal framework of statistical decision theory and applying the modern statistical methods. In Section 2, we describe our decision-theoretic framework which is based on the recently developed literature about treatment choice (e.g., Manski, 2005). We differ from this literature by adopting a Bayesian approach rather than a frequentist approach based on the maximin or minimax-regret criteria (see Dehejia, 2005, for a notable exception). We then show how the standard linear programming algorithm can be used to derive the optimal campaign strategy within this framework. Another methodological contribution of this paper is to address the subgroup analysis problem mentioned above, which has been somewhat neglected in the treatment choice literature. We propose a new variable selection algorithm and use it in combination with nonparametric methods and cross validation procedures in order to avoid over-fitting of statistical models (Section 3). These methods are then extended to the cases of optimal partisan campaign planning (Section 4). Together, we offer a set of methods that can be used to derive the optimal campaign strategy from randomized field experiments.

To illustrate our proposed methodology, we apply it to three existing field experiments evaluating the efficacy of nonpartisan text messages to mobilize voters, the impact of social pressure on turnout, and a partisan mobilization campaign (Section 5). In multiple cases, we find that the resulting optimal campaign strategy is more than twice as cost-effective than a naive strategy.

2 The Formal Framework of GOTV Campaign Planning

In this section, we follow the literature on treatment choice (e.g., Manski, 2005) and formalize the problem of deriving the optimal nonpartisan GOTV campaign strategy from experimental data as a statistical decision problem. We extend this framework to the optimal partisan persuasion and GOTV campaign in Section 4. Rather than taking a frequentist approach which is dominant in the treatment choice literature, we adopt the Bayesian statistical decision theory (Berger, 1985) and assume that a GOTV campaign planner learn about the effectiveness of various mobilization methods for different voters by analyzing randomized field experiments. The planner then maximizes the posterior expected turnout among a target population of voters subject to the budget constraint.[1]

2.1 The Planner's Decision Problem

We begin our analysis by formally stating the decision problem faced by a nonpartisan GOTV campaign planner. Let $\mathcal{P}$ denote this planner's target population of voters where this population

[1] In what follows, we describe our methodology as if a campaign planner both analyzes the data and makes the decision. Clearly, this does not have to be the case since our methods apply directly to the situations where a planner makes the decision based on the recommendations given by a data analyst.

is assumed to be finite and of size N. Typically, the target population is the registered voters in the electoral district, whose complete list is available to the planner (at least in the U.S.). Then, the planner's decision problem is to assign one of K available mobilization techniques (i.e., treatments) to each member i of this population. If a GOTV campaign may be planned at the levels of households or precincts, then i indexes appropriate units rather than individual voters. Although we assume the units of analysis are individual voters for notational simplicity, our proposed methodology is directly applicable to aggregate units.

We use an unordered set $\mathcal{T} \equiv \{0, 1, \ldots, K-1\}$ where $K \geq 2$ to denote the finite set of mobilization techniques from which the planner makes a selection for each voter in $\mathcal{P}$. Note that $T = 0$ represents the strategy of not mobilizing (i.e., doing nothing). For example, the planner may consider three strategies (i.e., $K = 3$) where $T = 2$ and $T = 1$ represent a GOTV method based on a phone call and a post card, respectively, and $T = 0$ denotes a status quo strategy that involves no such phone call or mailing. Alternatively, mobilization techniques may differ in their frequency, timing, contents of messages, and other aspects.

Next, suppose that the planner observes the J-dimensional (pre-treatment) covariates X for each member of this population $\mathcal{P}$. Since $\mathcal{P}$ is a finite population, this means that the planner knows the population distribution of X, i.e., $P(X)$, whose support is denoted by $\mathcal{X}$. For example, if the target list of voters is obtained from the voter roll, such covariates may include age, gender, voting history, party registration, and the zip code. Since the values of these covariates are observable for every voter on the list, the distribution of X is known to the planner.

Finally, following the statistical literature on causal inference, let $Y_i(t)$ represent the potential turnout of voter i that would be realized if the planner applies the mobilization technique t to this voter where $i \in \mathcal{P}$ and $t \in \mathcal{T}$. Then, the observed turnout is given by $Y_i = Y_i(T_i)$. The outcome variable is binary and is equal to 1 if voter i casts a ballot and is equal to 0 if the voter abstains. For the sake of simplicity, we assume no interference between voters (Cox, 1958; Rubin, 1990). But, this assumption can be easily relaxed. If a voter's turnout decision depends on the treatment status of the other voters within the same household (Nickerson, 2008), for example, then an analysis needs to be conducted at the household level.

Given this setup, the planner's mobilization strategy is characterized by the following function,

$$\delta(\cdot,\cdot) : (\mathcal{T} \times \mathcal{X}) \mapsto [0,1], \tag{1}$$

where the *mobilization strategy* $\delta(t,x)$ denotes the probability of receiving treatment $t \in \mathcal{T}$ for a voter with $X = x$. Alternatively, $\delta(t,x)$ may represents the fraction of voters with the observed covariate $X = x$ are contacted by the planner using the mobilization method t. These two definitions become essentially identical when the number of voters is large, but have different implications for the way the optimization is conducted. In either case, voters with the same value of X are assumed to be exchangeable since the planner does not possess additional information to distinguish them. Thus, the planner is interested only in determining the value of the function $\delta(t,x)$ for each value of $t \in \mathcal{T}$ and $x \in \mathcal{X}$. Given this definition, the set of feasible mobilization strategies, Δ, is the collection of functions $\delta(\cdot,\cdot)$ that satisfy the following complement property constraint,

$$\Delta \equiv \left\{ \delta(\cdot,\cdot) : \sum_{t=0}^{K-1} \delta(t,x) = 1 \text{ for every } x \in \mathcal{X} \right\}. \tag{2}$$

Now, the goal of a nonpartisan GOTV campaign planner is to derive the optimal mobilization strategy so that the turnout of the target population is maximized. The planner may achieve this by deriving the strategy that maximizes the expected turnout given the observed covariate information about X. Then, the planner's objective function can be written as a function of the planner's mobilization strategy as well as the probability of a voter's turnout given the values of covariates and the actual mobilization strategy applied to the voter,

$$g(\delta,\rho) \equiv \mathbb{E}\left(\sum_{i=1}^{N} \delta(t,X_i) Y_i(t) \;\middle|\; X \right) = N \sum_{x \in \mathcal{X}} P(X = x) \sum_{t=0}^{K-1} \delta(t,x)\, \rho(t,x), \tag{3}$$

where the *turnout profile* is denoted by $\rho(\cdot,\cdot) : (\mathcal{T} \times \mathcal{X}) \mapsto [0,1]$ with $\rho(t,x) \equiv \Pr(Y(t) = 1 \mid X = x)$. The turnout profile represents the turnout probability given the characteristics of a voter and the mobilization strategy applied to this voter. Note that the function $\rho(\cdot,\cdot)$ is unknown to the planner. If the turnout profile were known (and there were no budget constraint), then the planner would apply the mobilization technique t that gives the highest value of the function $\rho(t,x)$ given each voter's covariate characteristics x. However, since $\rho(\cdot,\cdot)$ is unknown in practice, the planner must make the decision under uncertainty by learning about $\rho(\cdot,\cdot)$ from experimental data.

A typical voter mobilization method usually has a small effect on an individual's turnout probability relative to the individual's baseline predisposition to vote. For this reason, planners may cast the objective function of equation 3 in terms of *treatment effect* rather than turnout profile. In this framework, the planner maximizes,

$$g(\delta,\rho) \;=\; \mathbb{E}\left(\sum_{i=1}^{N} \delta(t, X_i)(\tau_i(t) + Y_i(0)) \;\Big|\; X\right) \;\propto\; N \sum_{x\in\mathcal{X}} P(X = x) \sum_{t=1}^{K-1} \delta(t,x)\,\tau(t,x), \tag{4}$$

where $\tau_i(t) \equiv Y_i(t) - Y_i(0)$ is the treatment effect of mobilization strategy t on voter i and $\tau(\cdot,\cdot) : (\mathcal{T},\mathcal{X}) \mapsto [-1,1]$ with $\tau(t,x) \equiv \Pr(Y(1) = 1 \mid X = x) - \Pr(Y(0) = 1 \mid X = x)$. This setup is mathematically equivalent to equation 3. The decision whether to maximize turnout profile (equation 3) or treatment effect (equation 4) hinges on whether the planner can more easily elicit a prior for the turnout profile or treatment effects (see Section 3.2).

Furthermore, in practice, the planner cannot maximize the expected turnout without considering the differing costs of various GOTV techniques. We operationalize this idea by assuming that the planner faces the following budget constraint,

$$\sum_{t=1}^{K-1} \mathbf{1}\left\{\sum_{x\in\mathcal{X}} \delta(t,x) \neq 0\right\} \kappa(t) + N \sum_{x\in\mathcal{X}} P(X = x) \sum_{t=1}^{K-1} \delta(t,x)\,\xi(t,x) \;\leq\; C, \tag{5}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and C is the fixed positive constant representing the maximum cost allowed for the GOTV campaign. In this formulation, there are two kinds of costs the planner needs to consider. The first is the fixed *overhead cost* denoted by the function, $\kappa(\cdot) : \mathcal{T} \mapsto [0, \infty)$ for each mobilization technique t. This cost is incurred so long as at least one voter is assigned to the mobilization method. The second component is the *cost per voter*, i.e., the cost of contacting each voter with covariate value $x \in \mathcal{X}$, which is represented by a known cost function $\xi(\cdot, \cdot) : (\mathcal{T} \times \mathcal{X}) \mapsto [0, \infty)$. Thus, the planner must determine which mobilization technique is most cost effective for different voters, and whether such differences are large enough to warrant use of multiple mobilization techniques. Finally, we note that since $t = 0$ corresponds to the status quo strategy of not mobilizing, both the overhead and per voter costs are zero for this strategy, i.e., $\xi(0, x) = \kappa(0) = 0$ for all $x \in \mathcal{X}$. This is why the summation in equation 5 is taken with all possible values of t except $t = 0$. For the other mobilization strategies, we assume that the cost per voter is positive, i.e., $\xi(t, x) > 0$ for all $x \in \mathcal{X}$ and $t > 0$, whereas the overhead cost is non-negative, i.e., $\kappa(t) \geq 0$ for all $t > 0$.

2.2 Data from a Randomized Field Experiment

Given the above decision-theoretic framework, the planner can analyze the data from a randomized field experiment to derive the optimal GOTV campaign. Here, we formalize the assumptions required to link a field experiment with a planner's decision problem. First, we assume that the experiment is conducted on a representative sample of size n taken from the target population $\mathcal{P}$. Without such an assumption, the planner must model the non-random sample selection mechanism in order to infer about $\mathcal{P}$ from the experiment. We observe the data denoted by $D = \{\widetilde{Y}_i, \widetilde{T}_i, \widetilde{X}_i\}_{i=1}^n$ where $\widetilde{Y}_i$ is the binary turnout variable, $\widetilde{T}_i \in \mathcal{T}$ is the treatment variable representing mobilization techniques, and $\widetilde{X}_i \in \mathcal{X}$ is the same set of covariates as before. The potential outcomes are denoted by $\widetilde{Y}_i(t)$ where $\widetilde{Y}_i = \widetilde{Y}_i(\widetilde{T}_i)$ for $t \in \mathcal{T}$.

Another key assumption required for the planner to apply the results of a randomized field experiment to the decision problem is that the joint distribution of potential outcomes and covariates do not change, i.e., $P(\widetilde{Y}(t), \widetilde{X}) = P(Y(t), X)$. The assumption may be violated if, for example, the election in which the experiment was conducted differs significantly from the election for which the planner is designing the GOTV campaign. Although in real world settings this assumption may hold only approximately, it is a necessary requirement for learning about the planner's decision problem from a field experiment.

If the above assumptions hold, the randomization of treatments in field experiments imply that the turnout profile, $\rho(\cdot, \cdot)$, is identified, i.e., $\rho(t, x) = P(\widetilde{Y}(t) = 1 \mid \widetilde{X} = x)$, for all $t \in \mathcal{T}$ and $x \in \mathcal{X}$. Although large sample identification results such as this one are important, in practice the planner must estimate $\rho(\cdot, \cdot)$ from a finite sample and make the decision under uncertainty. (Similarly, we can identify $\tau(t, x)$ but must estimate it from a finite sample). Next, we address this problem by showing how to derive the optimal nonpartisan GOTV campaign strategy in this setting when the planner is assumed to be Bayesian.

2.3 The Bayesian Planner

We assume that the planner is Bayesian and has a prior belief on the space of functions of $\rho(\cdot, \cdot)$. This prior distribution is denoted by $\pi(\rho)$. The Bayesian planner will update her belief via Bayes rule after observing the data from the randomized field experiment described above. We use $\pi(\rho \mid D)$ to represent this posterior belief about the turnout profile. In the Bayesian statistical decision framework (e.g., Berger, 1985), the optimal nonpartisan GOTV strategy δ^* maximizes the posterior mean of the expected turnout, i.e.,

$$\delta^* \quad = \quad \underset{\delta \in \Delta}{\operatorname{argmax}} \int g(\delta, \rho) \; d\pi(\rho \mid D), \tag{6}$$

where the optimization is subject to the budget constraint given in equation 5.

There are several reasons why we use the Bayesian optimality criteria rather than a frequentist approach based on maximin or minimax-regret criteria that is popular in the treatment choice literature. First, the Bayesian decision has a frequentist justification. To see this, consider an alternative class of GOTV mobilization strategies which directly depend on experimental data as well as the values of observed covariates. Such strategies are called “statistical treatment rules” and are characterized by the function $\delta(\cdot,\cdot,\cdot) : (\mathcal{T} \times \mathcal{X} \times \mathcal{D}) \mapsto [0, 1]$ (Manski, 2005). Thus, the set of feasible strategies Δ equals the set of all functions $\delta(\cdot,\cdot,\cdot)$ that satisfy $\sum_{t=0}^{K-1} \delta(t, x, D) = 1$ for all $(x, D) \in (\mathcal{X} \times \mathcal{D})$. Under this setting, the frequentist objective function (i.e., risk) is given by,

$$g(\delta, F_D, \rho) \quad = \quad \int_D N \sum_{x\in\mathcal{X}} P(X = x) \sum_{t=0}^{K-1} \delta(t, x, D)\, \rho(t, x)\; dF_D \tag{7}$$

with the following budget constraint,

$$\sum_{t=1}^{K-1} 1\left\{\sum_{x\in\mathcal{X}} \delta(t, x, D) \neq 0\right\} \kappa(t) + N \sum_{x\in\mathcal{X}} P(X = x) \sum_{t=1}^{K-1} \delta(t, x, D)\, \xi(t, x) \quad \leq \quad C, \tag{8}$$

for all $D \in \mathcal{D}$. Again, the costs for the status quo strategy $t = 0$ are zero and excluded from the budget constraint. It has been shown that in most practically relevant situations the Bayesian decision δ^* defined in equation 6 agrees with the decision that maximizes the expected value of $g(\delta, F_D, \rho)$ averaging over the prior distribution of D on $\mathcal{D}$ (see Berger, 1985, p.159; Manski, 2005, p.59). Also, if the size of experimental data is large (as in many GOTV randomized field experiments) and little prior information is available, the Bayesian decision is essentially equivalent to the strategy that maximizes the expected turnout. Thus, the Bayesian decision can be justified from a frequentist perspective.

Second, an alternative optimality criterion which attracted attention in the literature is the minimax regret principle of Wald (1950) (see Savage, 1951; Manski, 2005). One important advantage of the minimax regret criterion is that it avoids the subjectivity of Bayesian optimality because it does not require the use of prior information. On the other hand, unlike the Bayesian decision theory, the frequentist theory based on the minimax regret criterion typically does not lead to the unique optimal decision, which practitioners may find problematic. In addition, the strategies that meet the minimax regret criterion can include no-data rules which do not depend on the data at all (Stoye, 2009). Clearly, such strategies are not desirable as they do not allow the planner to learn anything from the available experimental data. Furthermore, a minimax regret rule can be viewed as a Bayes rule with a prior (i.e., a least favorable prior) distribution (Berger, 1985, Chapter 5). Thus, depending on the planner's subjective belief, the Bayesian GOTV strategy can meet the minimax regret criterion.

2.4 Bayesian Optimal Campaign Planning at A Glance

Finally, we briefly summarize the proposed Bayesian decision theoretic framework. Figure 1 depicts the process by which a planner arrives at the optimal strategy. The planner must determine the costs of each mobilization strategy (both overhead and per voter) as well as the prior belief about their effects on voters with different characteristics. Since the cost function inputs are often exogenously determined (e.g., the cost of postage and phone calls), the planner's only meaningful decision might be determining a prior belief. In many cases, the planner might use a diffuse prior centered around a belief that there is no *a priori* difference in effects of a mobilization method across different voters. This is especially appropriate if the mobilization technique has not been previously empirically tested (e.g., airplanes with reminder to vote advertisements). If the treatment is an oft-used mobilization technique that is extensively studied in the past (e.g., canvassing), then the planner might center a prior around the estimated effects in previous experiments. Although the formalization of subjective belief as a form of probability distribution is a difficult topic and is beyond the scope of this paper, we note that the influence of prior belief diminishes as the size of experimental data increases.

Once a prior belief is elicited and a randomized field experiment is conducted, these two sources of information are combined via Bayes Rule to obtain the posterior belief about the effects of mobilization strategies on different voters. For every voter, the planner has a updated belief about

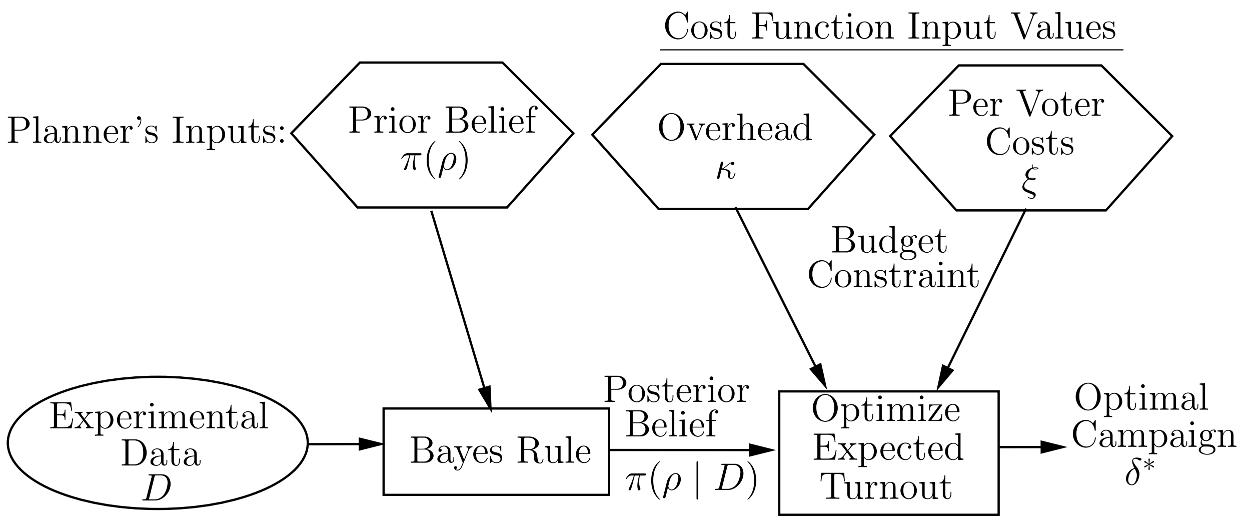

Figure 1: An Overview of the Bayesian Optimal Campaign Planning Process. Inputs over which the planner has direct control are represented by hexagons and are: (1) the prior belief about the effects of various mobilization strategies on different voters, $\pi(\rho)$, (2) the overhead costs of each mobilization method, κ, and (3) the cost per voter for each strategy, ξ. Data from randomized field experiments, D, are represented by the oval. This data and the planner's prior distribution are combined via Bayes Rule to produce a posterior belief about the effects of mobilization strategies $\pi(\rho \mid D)$. Finally, our proposed optimization method uses this posterior belief and the exogenous costs ξ to find the optimal campaign strategy δ^* for the planner.

the most cost effective way to mobilize that voter. Taking into account overhead costs, the optimal strategy may be to implement a subset of the available mobilization techniques, even if every technique is marginally optimal for at least one voter.

3 The Optimal Nonpartisan Campaign Strategy

Given the decision theoretic framework described in the previous section, we next show how to derive the optimal GOTV campaign strategy defined in equation 6. One relative advantage of the proposed Bayesian framework over a frequentist's approach is that we can completely separate the derivation of the optimal campaign strategy from the analysis of experimental data. That is, we first use statistical models to obtain the posterior belief of the turnout profile, $\rho(\cdot, \cdot)$. Conditional on this posterior turnout profile, we obtain the optimal campaign strategy by solving an optimization problem subject to a budget constraint.

3.1 The Optimization Method

Before describing our method to obtain the posterior turnout profile, we show how to obtain the optimal nonpartisan GOTV campaign strategy given the posterior turnout profile. Let $\tilde{\rho}(t, x)$ be the posterior turnout profile for each $t \in \mathcal{T}$ and $x \in \mathcal{X}$. Then, the optimal campaign strategy can be obtained by solving the following constrained optimization problem,

$$\delta^* \quad = \quad \underset{\delta \in \Delta}{\operatorname{argmax}}\, N \sum_{x \in \mathcal{X}} P(X = x) \sum_{t=0}^{K-1} \delta(t, x)\ \tilde{\rho}(t, x), \tag{9}$$

subject to the budget constraint given in equation 5.

To solve this optimization problem, we first consider the case of no budget constraint. In this case, the solution is trivial because one simply needs to use the most effective strategy for each strata defined by X. That is, the optimal campaign strategy is given by,

$$\delta^*(t, x) \quad = \quad \begin{cases} 1 \quad \text{if} \quad t = \operatorname{argmax}_{s \in \mathcal{T}} \tilde{\rho}(s, x), \\ 0 \quad \text{otherwise}, \end{cases} \tag{10}$$

for any given $x \in \mathcal{X}$. However, in many cases, the budget constraint may prevent the planner from employing campaign strategy δ^* defined in equation 10.

Next, consider the case where the budget constraint is binding so that the strategy defined by equation 10 is not feasible and yet there is no overhead cost. Unfortunately, in this case, the derivation of the optimal strategy is no longer trivial. Thus, as a general strategy, we use the fact that this is the following constrained linear optimization problem,

$$\delta^* \quad = \quad \underset{\delta \in \Delta}{\operatorname{argmax}}\, N \sum_{x \in \mathcal{X}} P(X = x) \sum_{t=0}^{K-1} \delta(t, x)\ \tilde{\rho}(t, x), \tag{11}$$

$$\text{subject to} \quad \begin{cases} \delta(t, x) \ \geq \ 0 \quad \text{for all } x \in \mathcal{X} \text{ and all } t \in \mathcal{T}, \\ \sum_{t=0}^{K-1} \delta(t, x) \ = \ 1 \quad \text{for all } x \in \mathcal{X}, \\ N \sum_{x \in \mathcal{X}} P(X = x) \sum_{t=1}^{K-1} \delta(t, x)\, \xi(t, x) \ \leq \ C. \end{cases} \tag{12}$$

Then, the standard linear programming algorithm can be used to obtain the optimal strategy, δ^*. If $\delta(t, x)$ represents the proportion of voters rather than the probability of treatment assignment, then this can be formulated as a mixed integer programming problem, which is more difficult but can be solved by applying an appropriate algorithm. Moreover, as the sample size increases, this difference vanishes.

Finally, when there exists overhead costs for the mobilization methods, we consider a case where only a subset of mobilization techniques are applied to at least one voter. In this case, we can solve the optimization problem in the same manner as the case without the overhead cost except that the total overhead cost is subtracted from the maximum budget allowed for the campaign, i.e., $C - \sum_{t=1}^{K-1} 1\left\{\sum_{x\in\mathcal{X}} \delta(t, x) \neq 0\right\} \kappa(t)$. The optimal strategy is obtained by considering all possible such subsets and their corresponding optimal strategy, and then choosing the strategy that yields the greatest posterior expected turnout.

Although the proposed general strategy based on linear programming is easy to implement, in some cases, an approximate method, which is more computationally efficient, may be preferred. In Appendix A, we consider a fast approximate solution to the non-partisan planner's optimization problem.

3.2 The Statistical Method

There are many statistical models that could be used to obtain the posterior distributions of the turnout profile. Nevertheless, when choosing a model, the following statistical issues need to be addressed. First, a statistical model needs to be flexible in order to avoid strong functional form assumptions. In particular, a nonparametric model is useful because a planner does not know the exact functional form of the turnout profile *a priori*. Second, since many of the covariates available in GOTV field experiments are discrete (e.g., party registration, previous turnout history), a statistical model should be able to accommodate discrete covariates. This excludes the use of certain binary nonparametric regression models that require covariates to be continuous. Third, the model fitting procedure should only require a minimal number of arbitrary decisions from a planner (or a data analyst who is working for the planner).

Finally, a model should not be overfitted to the data at hand and thus an appropriate set of pre-treatment covariates must be carefully selected. Overfitting tends to yield a model with poor performance in the actual election to which the derived mobilization strategy will be applied. This is the main problem of the post-hoc subgroup analysis as described in Section 1. In practice, this consideration is important because the sample size may not be large enough to accommodate a large number of pre-treatment covariates that are potentially useful for deriving the optimal campaign strategy. Thus, an appropriate variable selection algorithm is needed as a part of the model selection procedure.

Moreover, as Gunter, Zhu, and Murphy (2007) points out, most existing variable selection algorithms are developed for improving prediction rather than decision making. They are closely related but not the same concepts. For example, a powerful predictor of the outcome, i.e., a *predictive variable*, is not necessarily useful for decision making if its effect on the outcome is constant between the treatment and control groups. Similarly, a variable that explains a significant portion of treatment effect heterogeneity, i.e., a *prescriptive variable*, may not be selected by standard variable selection procedures if it does not predict the observable outcome (rather than the potential outcome) as well as other variables.

Below, we describe a method that meets the above criteria relatively well. Our method is similar to the one proposed by Gunter *et al.* (2007) but differs from it in that we use a tree-based method (Breiman *et al.*, 1984) rather than Lasso (in part because our outcome variable is categorical) and our measure of importance for prescriptive variables is somewhat different. (It is also possible to use Bayesian regression tree models (Chipman *et al.*, 2008; Hill and McCulloch, 2008)). To introduce our proposed method, we begin by describing a simple Bayesian approach to model the turnout among voters with the same characteristics of covariates via the binomial distribution. For the moment, we assume that the sample size is sufficiently large and thus there is no need for variable selection. Using a conjugate prior, we can write the model as,

$$W_{tx} \mid T = t, X = x \quad \sim \quad \texttt{Binom}(n_{tx},\ \rho(t,x)), \tag{13}$$

$$\rho(t,x) \mid X = x \quad \sim \quad \texttt{Beta}(a_{tx},\ b_{tx}), \tag{14}$$

for each $t \in \mathcal{T}$ and $x \in \mathcal{X}$ where W_{tx} is the number of voters with $T_i = t$ and $X_i = x$ who turned out, n_{tx} is the total number of such voters, and (a_{tx}, b_{tx}) are the prior parameters. It is well known that this model yields the following posterior distribution,

$$\rho(t,x) \mid Y_i, X_i = x \quad \sim \quad \texttt{Beta}(W_{tx} + a_{tx},\ n_{tx} - W_{tx} + b_{tx}), \tag{15}$$

where the posterior mean of $\rho(t,x)$ is given by $(W_{tx} + a_{tx})/(n_{tx} + a_{tx} + b_{tx})$.[2]

Although its simplicity is attractive, the above model is unlikely to work well in practice if the sample size is small relative to the number of unique values the observable covariates X take. In particular, if one conditions upon irrelevant covariates, then the sample size within each subgroup will be too small to yield informative inferences about $\rho(t,x)$. Such overfitting will then necessarily lead to a mobilization strategy whose performance in the actual election will be poor. On the other hand, if important covariates are not used to define subgroups, the planner will fail to differentiate across voters and will choose a suboptimal campaign strategy. Thus, as explained earlier, in most practical cases we seek a principled way to select relevant variables and form subgroups before applying the above standard Bayesian model.

Our proposed solution to this problem consists the following three steps. First, a variable selection algorithm is applied in order to decide which variable needs to be conditioned upon

[2]Alternatively, the planner may formalize a prior belief in terms of treatment effects. If a normal prior distribution is used, then the posterior of $\tau(t,x)$ can also be approximated by a normal distribution.

when deriving the optimal campaign strategy. Next, given the selected variables, a tree-based classification method (Breiman *et al.*, 1984) is fitted to each treatment/control group in order to identify relevant subgroups within the group.[3] We use a cross validation procedure to avoid overfitting. Finally, once all subgroups are identified in this way, we then simply apply the above Binomial-Beta model within each subgroup to obtain the posterior distribution of $\rho(t, x)$ for all t and x. We call this approach semi-Bayesian because the data are used twice — to form subgroups and then calculate the posterior distribution. Below, we present the details of our proposed method.

Step 1: (Selection of Predictive Variables) Fit a classification tree to the entire sample using all pre-treatment covariates and the treatment variable. Use K-fold cross validation on the misclassification rate to determine the value of the parameter which controls the complexity of the tree, e.g., the complexity parameter in `rpart()` implementation in R (Ripley, 1996, Chapter 7). Denote the predictive variables that are used in the final model by V, i.e., $V \subset X$.

Step 2: (Importance of Prescriptive Variables) Order each pre-treatment covariate, X_j for $j = 1, 2, \ldots, J$, based on the statistic, $S_j \equiv g^*_{j1} - g^*_{j0}$, where g^*_{jk} is the optimal overall turnout using the turnout profile $\hat{\rho}_{jk}(t, x)$ for $k = 0, 1$. We obtain $\hat{\rho}_{j1}(t, x)$ by fitting a classification trees for the treated subset of the data (i.e., $\widetilde{T}_i \geq 1$) and using V and X_j as covariates. Similarly, $\hat{\rho}_{j0}(t, x)$ is obtained by fitting a classification tree on the untreated subset of the date (i.e., $\widetilde{T}_i = 0$). Note that when fitting these trees, we use the value of the complexity parameter from Step 1.

Step 3: (Model Fitting) For each $j = 1, \ldots, J$ with $S_j > 0$, conduct the following:

(a) Randomly divide the sample into K subsets for K-fold cross validation.

(b) Using $K - 1$ training sets, fit classification trees (without pruning) separately to the treatment and control groups using V and j most important prescriptive variables in both models. Select the values of the complexity parameters for the two models based on the mean of the optimal overall turnout across K validation sets. Denote the optimal overall turnout and the optimal campaign strategy corresponding to the selected values of complexity parameters by g^*_j and δ^*_j, respectively.

Step 4: (Derivation of Optimal Strategy) The optimal overall turnout is given by $g^* = \max_{1 \leq j \leq J} g^*_j$, whereas the optimal campaign strategy is given by $\delta^* = \delta^*_{\operatorname{argmax}_{1 \leq j \leq J} g^*_j}$.

Each step of the above algorithm can be understood as follows. The first step selects predictive

[3] This tree-based classification method is one of many nonparametric models one can use. One disadvantage of this approach is that graduate changes in treatment effects across covariate groups are modeled as sharp discontinuities rather than smooth functionals.

variables using a standard fitting procedure of tree-based methods. The second step orders each pre-treatment covariate (including the ones identified as predictive variables in Step 1) according to its importance as a prescriptive variable. The statistic, S_j, is designed to measure how much a planner can increase the optimal overall turnout by interacting the value of X_j with treatment. This statistic provides a measure of the ability of X_j to explain heterogeneous treatment effects. The third step uses the K-fold cross validation procedure, given all predictive variables and different subsets of prescriptive variables, in order to select the values of complexity parameters for classification trees fitted separately to the treatment and control groups. This is done by comparing the optimal overall turnout corresponding to different values of the complexity parameters. Finally, the fourth step selects the final model among the ones chosen in Step 3 by again comparing the resulting optimal overall turnout and thus determines the optimal campaign strategy.

A main advantage of this semi-Bayesian approach is that it inherits the simplicity of tree-based methods. In particular, each of the subgroups that result from the final model is interpretable by practitioners. This means that it is possible to use available prior information within subgroups by specifying the parameters of the beta prior distribution. The proposed approach also addresses two key issues mentioned at the beginning of this section. The tree-based classification models, which our method uses, are nonparametric and can handle discrete covariates in an effective manner. The use of cross validation procedure avoids overfitting. Finally, transparent algorithms such as the one proposed here prevent planners from making arbitrary decisions when deriving the optimal campaign strategy.

4 The Optimal Partisan Campaign Strategy

In this section, we extend the proposed decision theory framework as well as the statistical and optimization methods described above to the case of a partisan persuasion and GOTV campaign planner. We show how a Bayesian planner can derive the optimal campaign strategy using randomized field experiments in order to maximize the (posterior) expected chance of winning the election. Throughout the section, we assume that two major candidates compete for the office. Minor party candidates may exist but it is assumed that their probability of winning the election is negligibly small.

4.1 The Decision Problem

Using the same notation introduced in Section 2, the decision problem of the partisan campaign planner is to assign one of K different mobilization methods (including the status quo strategy of doing nothing, which is denoted by $T_i = 0$) to each member of the target population $\mathcal{P}$ of finite size N. Again, we assume that the planner knows the distribution of a certain set of covariates $P(X)$. Thus, the planner's mobilization strategy can be characterized by $\delta(\cdot, \cdot)$ in the same way as before (see equation 1) and the set of feasible such strategies is equal to Δ which is defined in equation 2.

Unlike a nonpartisan campaign planner, however, a partisan campaign planner wishes to choose the mobilization strategy that leads to the election victory. For this decision problem, the outcome variable Y_i needs to be redefined in the following manner. Let $Y_i(t)$ represent the potential voting behavior of voter i that would be realized if the planner assigns mobilization method t to this voter where $i \in \mathcal{P}$ and $t \in \mathcal{T}$. The variable $Y_i(t)$ can take three different values; it equals 1 if voter i casts a ballot for the candidate of the planner's party, -1 if she votes for the opponent, and 0 otherwise (voting for a third party candidate or abstaining etc.). Then, the planner's ultimate goal is to win the election, which can be represented as the following indicator function,

$$h(\delta, V) \equiv 1\left\{\sum_{i=1}^{n} \delta(t, X_i) Y_i(t) > 0\right\} = 1\left\{\sum_{x \in X} P(X = x) \sum_{t=0}^{K-1} \delta(t, x) \nu(t, x) > 0\right\}, \quad (16)$$

where $\nu(t, x) \equiv \sum_{i \in \{i' : X_{i'} = x\}} Y_i(t) / \sum_{i=1}^{N} 1\{X_i = x\}$ is a random variable representing the *vote share differential* for the candidate that would result among voters with covariates $X = x$ if the planner assigns mobilization method t to them. Clearly, $h(\delta, V)$ is equal to 1 if the candidate of the planner's party wins the election, and it is equal to 0 if he/she loses. In the statistical decision theory literature, such an objective function is called "$0-1$ loss function."

Finally, the partisan planner typically faces a budget constraint similar to the one confronted by the nonpartisan planner, and therefore we use equation 5 for the partisan planner.

4.2 The Data Requirements

As is the case for a nonpartisan GOTV campaign planner, the assumptions are required in order for a partisan campaign planner to be able to directly use randomized field experiments for the optimal decision. These assumptions are essentially identical to those described in Section 2.2. That is, we require that (1) a field experiment is conducted on a representative sample from the same target population of voters $\mathcal{P}$, (2) the joint distribution of potential outcomes and covariates, $P(Y(t), X)$, remains identical between the experiment and the actual election. However, one important difference is that the derivation of the optimal partisan campaign requires vote choice data as well as turnout data for the voters who are subjects of field experiments; recall that in the case of partisan campaign $Y(t)$ represents a trichotomous variable rather than a binary variable. For example, in the United States, unlike turnout data, vote choice data are not publicly available and cannot be verified for each voter. This means that a sample survey needs to be conducted in order to derive the optimal partisan mobilization strategy unless the entire analysis and strategy planning are conducted at an aggregate level where validated election results are available.

4.3 The Derivation of the Optimal Strategy

Using the Bayes theorem, we derive the optimal partisan campaign strategy just as done in the case of nonpartisan GOTV campaign. In particular, the optimal strategy maximizes the posterior probability of winning the election,

$$\delta^* = \operatorname*{argmax}_{\delta \in \Delta} \int h(\delta, \nu) \; d\pi(\nu \mid D), \tag{17}$$

subject to the budget constraint given in equation 5 where $\pi(\nu \mid D)$ is the posterior distribution of the vote share differential $\nu(\cdot, \cdot)$. Using the classification method and variable selection algorithm similar to those described in Section 3.2, we can estimate $\nu(\cdot, \cdot)$ except that the outcome variable is now trichotomous rather than binary. Most existing classifiers including tree-based methods can handle such categorical variables even when the number of categories is greater than two.

The problem, however, is that the optimization in equation 17 is not trivial for two reasons. First, the integration cannot be explicitly solved. Second, the objective function is an indicator function which is not continuous. These difficulties are amplified by the fact that the optimization must be conducted over a high-dimensional space if the number of treatments and/or the number of subclasses are large. These computational considerations prevented Kramer (1966) from using the probability of winning as the objective function of a partisan campaign planner. Instead, he used the expected plurality of votes as the objective function while acknowledging that it may not be appropriate. Kramer (1966) explained this dilemma as follows,

> the probabilistic objective is the more realistic. However, this formulation is computationally quite difficult to work with (p. 141).

Indeed, if the partisan planner wishes to maximize the expected plurality, one can solve the following constrained optimization problem by applying the standard linear programming algorithm as explained in Section 3.1,

$$\delta^* = \underset{\delta \in \Delta}{\operatorname{argmax}} N \sum_{x \in \mathcal{X}} P(X = x) \sum_{t=0}^{K-1} \delta(t, x) \mathbb{E}(\nu(t, x) \mid D), \tag{18}$$

$$\text{subject to} \quad \begin{cases} \delta(t, x) \ \geq \ 0 \quad \text{for all } x \in \mathcal{X} \text{ and all } t \in \mathcal{T}, \\ \sum_{t=0}^{K-1} \delta(t, x) \ = \ 1 \quad \text{for all } x \in \mathcal{X}, \\ N \sum_{x \in \mathcal{X}} P(X = x) \sum_{t=1}^{K-1} \delta(t, x) \, \xi(t, x) \ \leq \ C. \end{cases} \tag{19}$$

To overcome this computational difficulty, we use a fast and approximate solution to the partisan planner's optimization problem to maximize the posterior probability of winning. This method, detailed in Appendix B, is not guaranteed to yield an optimal campaign strategy in some situations, but is relatively fast and approximates an optimal strategy.

5 Empirical Evaluation of the Proposed Method

How effective is the proposed method of deriving the optimal GOTV campaign work in real world applications? We empirically evaluate the performance of the proposed methods by analyzing three existing data sets of GOTV randomized field experiments. The basic idea here is to use a randomly selected subset of the data as a test data set and obtain an unbiased estimate of the actual turnout (or the probability of winning) by applying the resulting optimal campaign strategy derived from the rest of the data to this test data set. This procedure mimics the real world situation by using the test data set as the actual election to which the optimal campaign strategy is applied. Since the treatment is randomized and the test data set is not used to derive the optimal strategy, the procedure results in an unbiased evaluation of empirical performance of the proposed methodology.

Our three applications vary in their scope and context in order to demonstrate the wide applicability of our method. The first application is a single-treatment, text messaging nonpartisan GOTV experiment from the 2006 Congressional election. With only one treatment, the decision for the planner is which voters to treat with text messaging. Our second application is a multitreatment, nonpartisan GOTV effort that used mailings to pressure citizens to vote in the August 2006 primary election in Michigan. Here, the planner must decide which voter to contact and which mobilization method to use when contacting each voter. Finally, the third application is a partisan study with a single mobilization method where the planner's goal is to maximize the probability of winning the election.

5.1 Evaluation Method

To assess the effectiveness of the proposed method, we add an additional level of cross-validation to the procedure of Section 3. The key idea is to cross-validate the whole procedure consisting of the three steps described in Section 3.2 and obtain an unbiased estimate of the resulting turnout under the optimal strategy from test data that are not used for the derivation of the strategy. That is, after randomly dividing the sample into L subsamples, we set aside one of them as a test set and apply our methodology to the rest of the data. The derived optimal strategy is then applied to the test set in order to obtain an unbiased estimate of the resulting overall turnout under this strategy. The random assignment of treatments and the random subsampling of the test set make the unbiased estimation possible. The entire procedure is repeated L times using each subset as a test set. Finally, the average value of L estimated optimal turnout rates is taken as an estimate of the turnout that would result under our proposed methodology.

Throughout these applications, we use a normal-normal conjugate prior for the treatment effect τ based on the setup defined in equation 4. The prior for each subgroup treatment effect distribution is Gaussian and centered on the estimated population average treatment effect. The value of the prior variance is chosen so that it increases in proportion to the per capita budget constraint. In each case, a grid search is implemented to approximate the optimal complexity parameter. A 10-fold cross-validation procedure is used for Step 3 of Section 3.2 to determine the optimality of each complexity parameter. In addition, the optimization problem is solved using the algorithms detailed in Appendix which yield approximate (but fast) solutions.

5.2 A Nonpartisan GOTV Campaign with a Single Mobilization Method

During the 2006 election, two nonpartisan organizations contributed the cell phone numbers of newly-registered individuals to an experiment that tested the efficacy of text messages to mobilize voters (see Dale and Strauss, 2008a, for details). The election was of moderate interest, with at least one gubernatorial or senatorial campaign on the ballot in most states. Subjects were included in the experiment when they registered to vote with a campus representative of the Student PIRGs or when they registered online with Working Assets. About 8,000 subjects nationwide were randomly assigned, with 50 percent probability, to either the treatment or control group.

The treatment group received a short text message the day before Election Day. An example text message reads "A friendly reminder that TOMORROW is Election Day. Democracy depends on citizens like you-so please vote! -PIRG." The text message appeals were varied slightly, but these differences are ignored for this analysis. Subjects were matched to the voter file using information on their registration forms. The outcome variable is dichotomous: one for having voted in 2006, and zero for not voting. The estimated average treatment effect, or more precisely the overall intent-to-treat effect, was 3.0 percentage points with a standard error of 1.1. Available covariates for our analysis include: gender, age, race, past voting history, logged county population density, and registering organization.

We follow the procedure outlined in Section 3 and derive the optimal campaign strategy using this experimental data set. The classification tree produced in Step 1 chose the following variables as having predictive power (i.e., ν): age, log of county density, registering organization, whether the subject had voted in a previous election, and gender. The prescriptive variables chosen in Step 2 are age, density, and hispanic (in decreasing order of S_j). Not all of these variables are included in the final classification tree produced, however, as searching the complexity parameter space often

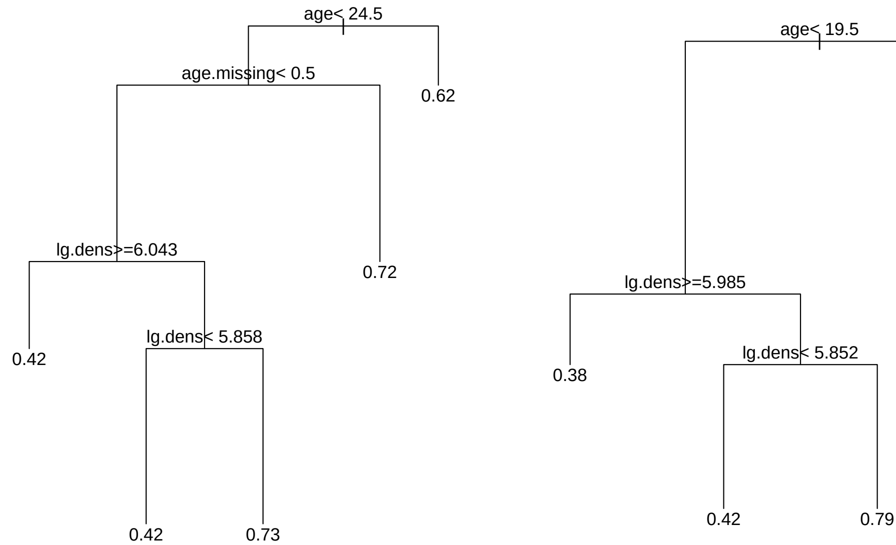

Figure 2: Final Classification Trees for the Control Group (left panel) and Treatment Group (right panel). The complexity parameters are chosen from 10-fold cross-validation using the algorithm described in Section 3.2 so that the resulting optimal turnout is maximized on the validation set. In this example, the planner budget allows treatment of at most 10% of the population. At each node, subjects who meet the node's criterion are filtered through the left branch of the tree. Covariate abbreviations: `age` is the age in years of the subject, `age.missing` is whether the age of the participant is unknown, and `lg.dens` is the log of the subject's county population density.

finds that models with fewer variables result in higher overall turnout on validation sets.

The resulting final trees are presented in Figure 2. The tree for the control group is on the left; the tree for the treatment group is on the right. Left branches represent voters who meet the criteria of the nodes; right branches represent voters who would falsify the nodes' inequalities. The leaves show the predicted probability of voting conditional on the leaves' covariates; higher probabilities are to the right at each node. The control group tree demonstrates that, in this group, voters above the age of 24 are predicted to vote at a rate of 62%, while whether the participant's age is known and county population density are important determinants of the voting rates of voters aged 24 or under. In the treatment group the age cut-point is in between 19 and 20, with population density and voting history also providing information about turnout rates.

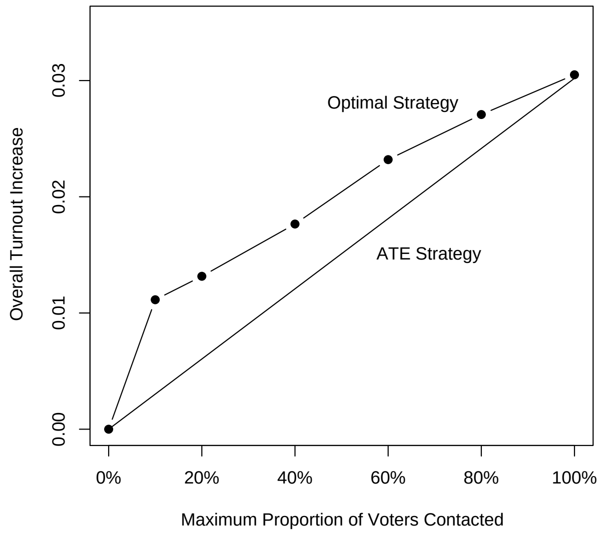

Figure 3: Empirical Evaluation of the Performance of the Proposed Method for the Text Messaging Experiment. The figure displays the overall turnout increase that results from two campaign strategies as a function of the maximum proportion of voters contacted. The first strategy is the average treatment effect or ATE strategy (solid line), which contacts randomly selected voters. The second strategy is an optimal approach based on the methodology outlined in this paper, which uses covariate characteristics of voters to determine which voters receive the treatment. Solid circles represent the estimated optimal turnout using the difference-in-means estimator. The estimator is applied to test data which are not used for the derivation of the optimal strategy.

Much of the heterogeneous treatment effect can be explained by the voters' ages. Potential voters between the ages of 20 and 24 are very responsive to the treatment. In the control group (left panel of Figure 2), these individuals vote at a rate of 62%. In contrast, the classification tree for the treatment group assigns them a probability of voting of 66% (the right branch of the right panel of Figure 2) – a four percentage point increase, above the average treatment effect. Also, the treatment model predicts a negligible or negative treatment effect for 18- and 19-year-olds, as their probability of voting is assigned at most 42% under the treatment unless they live in a county with a density inside a narrow range. Note that age-ranges such as these could not be identified by a classic model such as logit with linear explanatory variables.

Figure 3 displays the performance of our proposed methodology given the maximum proportion of voters that could be contacted. The turnout that would result under our optimal strategy is estimated using the difference in means estimator between the treated and untreated voters (solid lines with solid circles). That is, we compute the average turnout among the treated voters who are assigned to the treatment group under the optimal strategy as well as the average turnout among the untreated voters who are assigned to the control group under the optimal strategy. The latter is then subtracted from the former to yield the estimated overall turnout under the optimal strategy. The turnout rate achieved with the proposed method compares favorably with the turnout based on a naive strategy where randomly selected voters are contacted. This less-informed strategy, which is solely based on the estimated overall average treatment effect ("ATE strategy"), completely ignores covariate information and thus assumes zero treatment effect heterogeneity.

Figure 3 shows that the application of our proposed method allows one to achieve a higher overall turnout than an ATE strategy (at least on average), if an organization can only afford to treat a subset of the population. For a campaign that can afford to treat 10% of the population, the optimal campaign strategy achieves an overall turnout on average over three times greater than the turnout increase under the ATE strategy. Because the nonparametric procedure did not isolate individuals who reacted negatively toward the treatment, no gain over the ATE strategy is present when treating the entire population. This lack of negatively-responsive individuals is unsurprising given that little backlash was found to the text messages in a post-treatment survey (Dale and Strauss, 2008b). Overall, for the text messaging application, organizations that desire to pinch pennies would be wise to use the proposed nonparametric procedure, while large-budget mobilization campaigns achieve less gain.

5.3 A Nonpartisan GOTV Campaign with Multiple Mobilization Methods

In the 2006 Michigan August primary, scholars used the public voting records to test whether social pressure of seeing your or your neighbors' vote records would encourage individuals to vote

(Gerber *et al.*, 2008). The effort was extremely successful: learning about your neighbors' vote history increased the probability of casting a ballot from 0.30 to 0.38. This eight percentage point effect from the black-and-white postcard treatment is an order of magnitude above the usual turnout effect of mailings.

Over 300,000 participants are involved in this experiment which had three treatments. The treatments can be ordered in an hierarchy when each successive treatment includes all the aspects of the prior treatments. The civic duty message is common to all treatments. The "Hawthorne treatment" lets individuals know they are being studied. The "Self treatment" implicitly acknowledges that the participant is being studied when it provides the participant's voting record. The "Neighbors treatment" provided information on both the participant's records *and* those of their neighbors, and found to be associated with the largest estimated average treatment effect.

The outcome variable is whether or not the participant voted in the 2006 primary election. Available covariates are age, gender, and voting history (primary election turnout for 2000, 2002, and 2004 and general election turnout for 2000, and 2002 – all participants voted in the November 2004 election by design). The treatment costs are considered equal (all were printed on black-and-white postcards and sent through the U.S. postal service). Thus, the Neighbors condition is the most efficient in terms of votes per dollar. The procedure of Section 3.1 with the exception that because of the large sample size, 10-fold cross validation is not performed; instead, a training set of 20,000 participants is applied to a (randomly selected) test set of 125,000 participants.

As shown in Figure 4, the data exhibit little heterogeneity of treatment response, at least for the given covariates. The dotted line represents the boost in turnout under the non-experimental strategy; i.e., the outcome one might expect if the planner does not run an experiment and attempts to mobilize randomly selected voters by randomly selected strategies. The solid line is the turnout boost that would result from treating individuals randomly using the mobilization strategy that

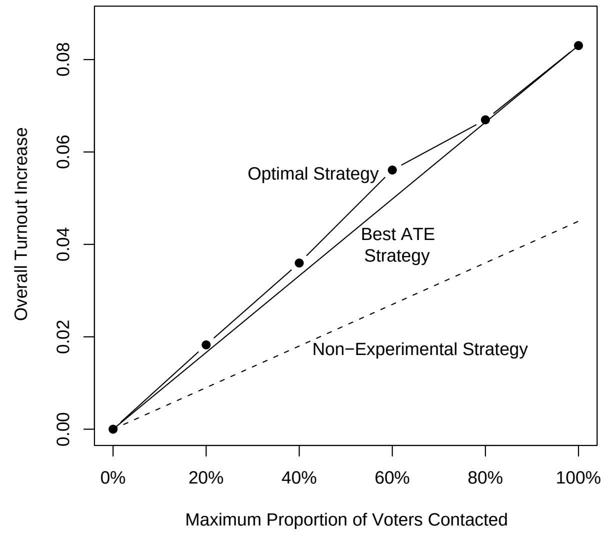

Figure 4: Empirical Evaluation of the Performance of the Proposed Method for Social Pressure Experiment. The figure displays the estimated overall turnout increase that results from three campaign strategies as a function of the maximum proportion of voters contacted. The first strategy (dashed line) is the non-experimental approach in which a mobilization method is chosen randomly and is applied to randomly selected voters ("Non-experimental Strategy"). The second strategy is the non-covariate experimental approach (solid line), which contacts randomly selected voters with the treatment that has the highest average effect ("Best ATE Strategy"). The third strategy is an optimal approach based on the methodology outlined in this paper, which uses covariate characteristics of voters to determine which voters receive the treatment. Solid circles represent the estimated optimal turnout using the difference-in-means estimator. The estimator is applied to (randomly selected) test data ($n = 125,000$) which are not used for the derivation of the optimal strategy (training set $n = 20,000$).

yields the best estimated average treatment effect; i.e., the outcome expected if the planner conducts an experiment but ignores covariates. The connected circles represent increase in turnout using the information in the covariates. In this example, we find that the covariates are not very helpful, though they do no harm. For each budget constraint we consider, all participants are optimally treated with the Neighbors treatment. This example demonstrates that using this method even when there is little heterogeneity only helps. Thus, even when a planner is unsure about the level of heterogeneity of the effect of the desired treatment, this procedure can be recommended.

5.4 A Partisan Persuasion Campaign

In September 2006, a "well-known liberal activist group" endorsed two Democratic candidates for the Pennsylvania State House and canvassed voters who they thought might be persuadable (see Arceneaux and Kolodny, 2007, for details). To assess the effectiveness of their message, randomly selected voters were canvassed. A post-election survey of 2,000 respondents recorded the vote preferences of the populations, of which about 30% received the personal door-to-door canvassing. Unfortunately, the original analysis of the experiment showed that the message miserably failed to persuade voters to support the Democrats. In fact, it ended up mobilizing Republican voters; the average (intention-to-treat) effect was estimated to be negative 6 percentage points (difference in vote margins) with the standard error of 4 percentage points. Arceneaux and Kolodny (2007) conjecture that the name of the liberal group in the endorsement provided useful information to Republican voters: since the disliked liberal group supported the Democratic candidate, the Republican voters increased their support for the Republican candidate.

To make the evaluation of our proposed method more efficient and realistic, we transform the data in two ways. First, we combine the two districts into one large district (i.e., voting for either Democrat counts as a vote for the Democrat). Second, since the treatment yielded a large and negative effect, we derive the optimal campaign strategy from the perspective of the Republicans. Indeed, an analysis from the Democratic perspective finds that the optimal strategy is to contact a very small percentage of voters. Available covariates we use are (from the voter file:) gender, age, (from the survey:) party identification, abortion views, senate vote, gubernatorial vote, and level of political interest. As before, we assume that the survey respondents are representative of the actual electorate.

Our algorithm finds two predictive variables (which also may be prescriptive), party identifi-

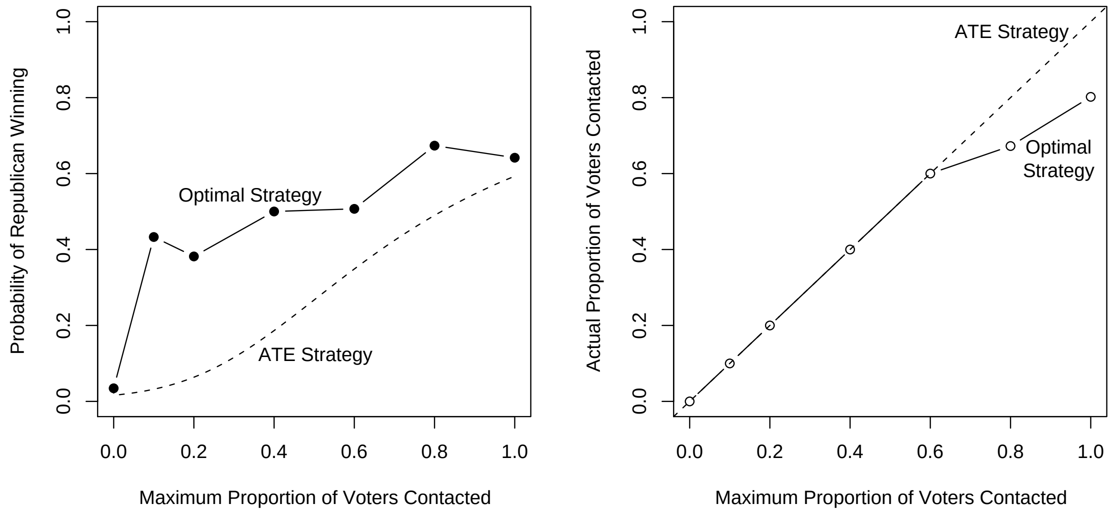

Figure 5: Empirical Evaluation of the Performance of the Proposed Method for the Partisan Example. The results are calculated based on 10-fold cross-validation. The left panel displays the estimated probability of the Republican candidate winning under two treatment strategies plotted over the maximum proportion of the electorate treatable under the budget constraint. The dashed lines represent the "ATE strategy" in which random voters are contacted. The solid lines represent the optimal strategy based on the methodology outlined in this paper, which uses covariate characteristics of voters to determine which voters receive the treatment. The optimal strategy outperforms the ATE approach strategy in most cases. The right panel plots the actual proportion of voters contacted by canvassing against the maximum proportion of voters contacted, which is determined by budget constraint. Only in the situations where nearly all voters can be contacted, does the algorithm choose not to canvass some voters.

cation and senate vote, and several purely prescriptive variables including: political knowledge, political interest, and views on abortion. The exact combination of the prescriptive variables depends on the budget constraint. This variable selection is in line with the analysis of Arceneaux and Kolodny (2007) which finds heterogeneous treatment effects among low-interest Republicans as well as differential effects depending on pro-choice/pro-life views. Figure 5 compares the optimal strategy with the ATE strategy in which randomly selected voters are contacted. The left panel shows that under the optimal strategy just contacting the 10% of the electorate increases the probability of Republican winning to about the same level as what could be achieved under the ATE strategy by contacting 70% of the electorate. The algorithm generally incrementally improves upon this result, when the budget is increased so that a maximum of between 20% and 60% of the electorate can be contacted. Finally, the right panel shows that even if the partisan planner had the resources to contact every voter, some of the electorate is not worth contacting. All together, the analysis shows that the proposed algorithm allows partisan GOTV campaign planners to efficiently increase the probability of winning with limited resources.

6 Concluding Remarks

More than forty years ago, Kramer (1966) started his pioneering article with the following observation,

> In the past two decades, the use of quantitative methods as aids for decision-making has become common in many fields. [...] By and large, however, these efforts have not been made by political scientists. [...] This is unfortunate, for many of the traditional concerns of political scientists appear to be quite susceptible to this sort of analysis. (p.137)

Despite this plea, political scientists have generally stayed away from the use of quantitative methods to directly inform policy makers who must make decisions using available data. In this paper, we take up Kramer's proposal and show how modern statistical methods can be used to help practitioners devise and implement the optimal policy.

We apply our method to the area of voter mobilization campaigns where political scientists have recently conducted over one hundred randomized field experiments. We demonstrate how to exploit these experiments and derive the optimal campaign strategy for both nonpartisan and partisan campaigns. We first use statistical decision theory to formalize the decision problem faced by GOTV campaign planners, and then propose a set of statistical and optimization methods that can be used to derive the optimal campaign strategy from experimental data. In the empirical evaluations, we find that the optimal strategy based on our proposed method is much more cost-effective than a naive strategy. As noted by Kramer, this paper shows that there is a great potential for political scientists to improve public policy through decision theoretic analysis.

Computational Appendix

In this appendix, we present fast and approximate solutions to the planner's optimization problem in both partisan and non-partisan cases.

A Nonpartisan Case: the Knapsack Problem

We first show how to approximate the solution to the non-partisan planner's optimization problem defined in Section 3.1. The key is to notice that the above linear optimization problem is identical to the canonical *knapsack problem*, in which one maximizes the total value of objects to be placed in a knapsack of fixed sized, with each object having its own value and size. The analogous case for the nonpartisan planner is maximizing the number of voters given a budget constraint where each individual-treatment pairing may be thought of as an object.

Following Dantzig (1957), we approximate the exact solution of this linear programming problem by ordering the individual pairs by their maximum vote per dollar ratio and treat the individuals with the highest such ratio first until the budget is exhausted. If the ratio is non-positive (i.e., the best non-control treatment for an individual does not outperform the control), this individual is not treated. In most cases, this approximation yields solutions very close to the optimal result because the ratio of the per-use cost of the most expensive treatment (e.g., \$15 for a canvassing shift) is tiny compared to the overall budget (usually at least \$10,000). Thus, when the addition of an expensive and efficient treatment runs just over budget and a cheaper yet less efficient tactic should be used in its place, inefficiencies at the edge of the problem are of little importance.

B Partisan Case: the Stochastic Knapsack Problem

To derive a fast and approximate solution to the partisan's optimization problem defined in Section 4.3, the key is to notice that this optimization problem is identical to the *stochastic Knapsack problem*, in which one maximizes the probability that the total value of items in the knapsack equals or exceeds a target value where each object has a random value and a known size. As in the non-partisan case, each individual-treatment pair can be treated as an item.

As an approximate solution to this problem, we use the algorithm that is based on Geoffrion (1967) where subgroups are ordered by the weighted combination of the mean and standard error of their posterior vote choice profile, $\pi(\rho)$. Optimization is performed over the weight parameter, which can take values between 1 (i.e., only mean of the posterior matter) and 0 (i.e., only the standard error matter). For a discussion of when this approximation fails to yield the optimal result, see Henig (1990). The intuition behind this algorithm can be developed by considering the following scenarios. Campaigns with a natural advantage (i.e., would garner a majority of the vote without treatment) could further increase their probability of winning by contacting voters who are highly responsive on average and have a low variance of their treatment response. On the other hand, campaigns who are behind aim to treat segments of the population who are *both* highly responsive and who have high variance. Thus, unlike in the nonpartisan case, the optimal subgroups to treat change depending on the outcome under the control. The algorithm finds an approximate solution by limiting its search to the subspace defined by the weight parameter, which makes optimization feasible when the dimension of δ is large.

References

Arceneaux, K. and Kolodny, R. (2007). The messenger or the message? group endorsements, heuristics, and grassroots campaigning. *Unpublished manuscript* .

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer, New York, 2nd edn.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. (1984). *Classification and Regression Trees.* Chapman & Hall, London.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2008). BART: Bayesian additive regression trees. Tech. rep., Department of Statistics, University of Pennsylvania. Available at `http://arxiv.org/abs/0806.3286v1`.

Cox, D. R. (1958). *Planning of Experiments.* John Wiley & Sons, New York.

Dale, A. and Strauss, A. (2008a). Don't forget to vote: Text message reminders as a mobilization tool. *Unpublished manuscript* .

Dale, A. and Strauss, A. (2008b). Mobilizing the mobiles: How text messaging can boost turnout. In C. Panagopoulos, ed., *Politicking Online: The Transformation of Election Campaign Communications.* Rutgers University Press, New Brunswick, NJ.

Dantzig, G. B. (1957). Discrete-variable extremum problems. *Operations Research* **5**, 2, 266–277.

Dehejia, R. (2005). Program evaluation as a decision problem. *Journal of Econometrics* **125**, 141–173.

Eldersveld, S. J. (1956). Experimental propaganda techniques and voting behavior. *American Political Science Review* **50**, 1, 154–165.

Geoffrion, A. M. (1967). Solving bicriterion mathematical programs. *Operations Research* **15**, 1, 39–54.

Gerber, A. S. (2004). Does campaign spending work?: Field experiments provide evidence and suggest new theory. *American Behavioral Scientist* **47**, 5, 541–574.

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

Gerber, A. S., Green, D. P., and Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review* **102**, 1, 33–48.

Gerber, A. S., Green, D. P., and Shachar, R. (2003). Voting may be habit-forming: Evidence from a randomized field experiment. *American Journal of Political Science* **47**, 3, 540–550.

Gosnell, H. F. (1927). *Getting-Out-the-Vote: An experiment in the stimulation of voting*. University of Chicago Press, Chicago.

Green, D. P. and Gerber, A. S. (2008). *Get Out the Vote: How to Increase Voter Turnout*. Brookings Institution, Washington D.C., 2nd edn.

Gunter, L., Zhu, J., and Murphy, S. (2007). Variable selection for optimal decision making. *Poceedings of the 11th Conference on Artificial Intelligence in Medicine* **AIME 2007, LNCS/LNAI 4594**, 149–154.

Hartmann, G. W. (1936). A field experiment on the comparative effectiveness of "emotional" and "rational" political leaflets in determining election results. *Journal of Abnormal Psychology* **31**, 1, 99–114.

Henig, M. I. (1990). Risk criteria in a stochastic knapsack problem. *Operations Research* **38**, 5, 820–825.

Hill, J. L. and McCulloch, R. E. (2008). Bayesian nonparametric modeling for causal inference. Tech. rep., Graduate School of Business, University of Chicago.

Kramer, G. (1966). A decision-theoretic analysis of a problem in political campaigning. In J. L. Bernd, ed., *Mathematical Applications in Political Science, Vol. II*, 137–160. Southern Methodist University, Dallas, TX.

Manski, C. F. (2005). *Social Choice with Partial Knowledge of Treatment Response*. Princeton University Press.

Nickerson, D. and Arceneaux, K. (2006). Who is mobilized to vote? a meta-analysis of 7 field experiments. In *Annual Meeting of the Midwest Political Science Association*.

Nickerson, D. W. (2007). Quality is job one: Professional and volunteer voter mobilization calls. *American Journal of Political Science* **51**, 2, 269–282.

Nickerson, D. W. (2008). Is voting contagious?: Evidence from two field experiments. *American Political Science Review* **102**, 1, 49–57.

Pocock, S. J., Assmann, S. E., Enos, L. E., and Kasten, L. E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine* **21**, 2917–2930.

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge.

Rubin, D. B. (1990). Comments on "On the application of probability theory to agricultural experiments. Essay on principles. Section 9" by J. Splawa-Neyman translated from the Polish and edited by D. M. Dabrowska and T. P. Speed. *Statistical Science* **5**, 472–480.

Savage, L. J. (1951). The theory of statistical decision. *Journal of the American Statistical Association* **46**, 253, 55–67.

Stoye, J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics* Forthcoming.

Wald, A. (1950). *Statistical Decision Functions.* Wiley, New York.