Randomized Evaluation of Institutions: Theory with Applications to Voting and Deliberation Experiments*

Yves Atchade† and Leonard Wantchekon‡

April 23, 2009

Abstract

We study causal inference in randomized experiments where the treatment is a decision making process or an institution such as voting, deliberation or decentralized governance. We provide a statistical framework for the estimation of the intrinsic effect of the institution. The proposed framework builds on a standard set-up for estimating causal effects in randomized experiments with noncompliance (Hirano-Imbens-Rubin-Zhou [2000]). We use the model to re-analyze the effect of deliberation on voting for programmatic platforms in Benin (Wantchekon [2008]), and the effect of direct democracy on public goods provision in Indonesia (Olken [2008]).

1 Introduction

Randomized experiments are a widely accepted approach to infer causal relations in statistics and social sciences. The idea dates back at least to Neyman (1923) and Fisher (1935) and has been extended by D. Rubin and coauthors (Rubin (1974), Rubin (1978), Rosenbaum and Rubin (1983)) to observational studies and other more general experimental designs. In this approach, causality is defined in terms of potential outcomes. The causal effect of a treatment, say Treatment 1 (compared to another treatment, Treatment 0) on the variable Y and on the statistical unit i is defined as $Yi(1) - Yi(0)$ (or its expected value) where $Yi(j)$ is the value we would observe on unit i if it receives Treatment j. The estimation of this effect is problematic because unit i cannot be given both Treatment 1 and Treatment 0. Randomizing the assignment of units to treatments allows us to overcome this difficulty. To estimate the causal effect of a treatment, two random samples of units are selected, the first group is assigned to Treatment 0 and the second group to

*Very preliminary and incomplete.

†Assistant Professor of Statistics, University of Michigan.

‡Professor of Politics and Economics, New York University.

Treatment 1. The difference in the sample means of Y (or some other statistic of interest) over the two groups is used as an estimate of the causal effect of the treatment. The main idea is that randomization eliminates (at least in theory) any systematic difference between the two samples. [1]

The past ten years have seen a sharp increase in the use of randomized experiments in development economics and political science. Researchers and policy makers became increasingly concerned about identification of the effects of programmes in face of complex and multiple channels of causality (Banerjee and Duflo [2008]. p. 2). Some early experiments in economics were interested in identifying the causal effects of various education inputs such as textbooks, and the student-teacher ratio on learning; Others looked at the effect of the treatment of intestinal worms on various measures education outcomes. Randomized field experiments in political science have primarily focused on studying the way in which various techniques of voter mobilization (mail, canvass, telephone) affect voter turnout.[2] More recent work covers a very wide range of topics such a women leadership, corruption, conditional cash transfer programmes, clientelist and programmatic politics. They also use increasingly refined and reliable identification strategies. (See Duflo (2008) and Gerber and Green (2007) for a survey).

In most previous research, the treatment is conceived and designed by the experimenter and assigned to an individual or a group of individuals. There might be compliance problems, i.e. individuals in active treatment groups might ex-post choose to enter the control group or vice-versa (see Imbens and Rubin (1997) and Angrist, Imbens and Rubin (1996)). It might also not be legally feasible to assign individuals to treatment or control groups, so the experimenter simply encourages individuals to take treatment 1 (and individuals so encouraged comprise the treatment group) (Hirano, Imbens, Rubin and Zhou (2000)). The policy to be evaluated might lack clarity or its implementation might be imperfect (Harrison, Lau and Rutström (2005)). In all these cases, there is a difference between the treatment assigned and the treatment received and this has been dealt with in a variety of ways by the encouragement design, non-compliance and treatment uncertainty literature.

Now assume that the treatment or the policy to be evaluated is an unknown outcome of a well specified process. That is, groups of individuals are randomly assigned to decision-making processes that allows them to pick the treatment they will eventually receive. For instance, instead of assigning schools to textbooks, flip charts or deworming treatments, we assign them a decision-making process over these three possible treatments,whereby parents and teachers use simple majority rule to decide whether all the classrooms should receive textbooks or flip charts,

[1] See Holland (1986) among others for a review.

[2] Gosnell (1927), Elderveld (1956), Adams and Smith (1980), Miller, Bositis and Baer (1981) and more recently Green and Gerber (2000).

or all the students should be treated with deworming drugs. Instead of majority rule, the decision-making process could be a strict proportionality rule: if α percent of the parents and teachers prefer X, then a proportion α of the school budget should be spent on X. This type of experiment would help identify the causal effect of either education input, when they are endogenously selected by parents and teachers. It could also help identify the intrinsic effect of majority or proportionality rule, and this result would have implications for evaluating not only education policies, but any other public policies. The study would also contribute to empirical studies of institutions by providing a rigorous test of the causal effect of majority and proportionality rule on a variety of outcomes.[3]

There are at least two recent papers that explicitly integrate institutions or decision-making processes in field experiments. Olken (2008) provides experimental evidence from Indonesia on the effect of direct democracy on support for public goods provision. The experiment involves 49 villages that were assigned to select development projects either through direct elections or meetings of local leaders. The author finds that direct participation has a positive effect on satisfaction among villagers, knowledge about the project and willingness to contribute, but finds no significant difference between direct democracy and representative-based meetings in terms of the project picked. In a paper using similar approach, Wantchekon (2008) provides experimental evidence on the combined effect of "informed" platforms and public deliberation on electoral support for programmatic, non-clientelist platforms. The experiment takes place in Benin and involves 5 candidates running in the first round of the 2006 presidential elections. The treatment to be evaluated is a two-stage public deliberation process. In the first stage, policy experts helped candidates design electoral platforms that are specific and transparent in terms policy promises. In the second stage (during the elections), there were town meetings in treatment villages, while there were rallies in control villages. The author finds that the treatment (specific platforms and town meetings) has a positive effect on voter information about policies and candidates. He finds that both turnout and electoral support for the candidate running the experiment was higher in treatment areas than in control areas (even though the turnout result was more significant than electoral support result).

One important limitation of these two papers is that they could only identify ITT type effects and did not really isolate the intrinsic effect of the institutions from the effect of the selected policy. For instance, in Olken (2008), satisfaction is higher under direct democracy than under representative meetings, but it is unclear if the result was driven by democracy or by the outcome of democracy (i.e. the project that was picked by plebiscite). In fact, they are number of cases where the projects selected under democracy were different from the ones

[3] The study would also be of great interest for policy-makers since it incorporates political economy considerations in the impact evaluation of education inputs on learning.

selected under representative meetings. Thus, the difference in satisfaction could well be driven at least in part by differences in policy selected under the two political mechanisms. In addition, since the selection of policy is endogenous, a simple comparison of groups that have selected the same policy under direct democracy and representative-based meeting can lead to a selection bias. As for Wantchekon (2008), treatment groups have town meetings where specific policies were discussed as opposed to control villages where rallies where held and mostly clientelist platforms were presented. The paper did not investigate wether the effect of the treatment was driven by the information content of the electoral platforms or by the institution of the town meetings. The goal of this paper is to provide a statistical model that disentangle these two effects thereby helping to identify the intrinsic effect of the institution.

In the next section, we will present the statistical framework. We then apply it to the town meeting experiment in Benin and the democracy experiment in Indonesia. Finally we provide general guidelines for the conduct of randomized evaluation of institutions.

2 The Model

2.1 Defining the causal effects

Suppose we have two decision-making processes or institutions denoted 0 and 1. The processes are assigned to communities, i.e. groups of individuals. For simplicity, we assume that Process 0 is the status-quo or the control, which consists of applying a clearly defined policy (called Treatment 0) to the community; whereas in Process 1, the community is given the possibility of choosing through some decision-making process (e.g. voting, deliberation) any treatment in a set $\{0, \ldots, L\}$. The Treatment 0 from that set is the same as the treatment applied under Process 0. Let Y an outcome variable of interest that will be measured after the Treatment is applied. Let Z be an indicator variable indicating which process is applied to the community. Let $D \in \{0, \ldots, L\}$ be the treatment choice made by the community under Process 1.

Let $Y(0)$ the potential outcome under (Treatment 0 of) Process 0. Let $Y(1, d)$ be the potential outcome of under Treatment $d \in D$ of Process 1. We can define the causal effect of Process 1 compared to Process 0 as

$$\tau_0 = \mathbb{E}\left(Y(1, D) - Y(0)\right). \tag{1}$$

The effect τ_0 corresponds to the overall effect of Process 1 versus Process 0 and includes both the effect of the selected policy D and the effect of the decision-making process. By encouraging people participation and exchange of information, the decision-making process itself can have a profound effect on people's satisfaction.

We also introduce the causal effect of Treatment d versus Treatment 0 (under Process 1):

$$\tau_{1,d} = \mathbb{E}\left(Y(1,d) - Y(1,0)\right).$$

This measures the intrinsic effect of policy d versus policy 0 under Process 1.

We can pool together these $\tau_{1,d}$ to define what we can the causal effect of Treatment (under Process 1) as

$$\tau_1 = \mathbb{E}\left(Y(1,D) - Y(1,0)\right). \tag{2}$$

Another quantity of interest is what we call the *intrinsic effect of the process* define as

$$\tau_2 = \mathbb{E}\left(Y(1,0) - Y(0)\right). \tag{3}$$

Clearly, the overall effect of Process 1 versus Process 0 can be written as $\tau_0 = \tau_1 + \tau_2$. The term τ_1 can be further decomposed in term of the effect $\tau_{1,d}$. This is done in the following Proposition.

Proposition 2.1. *We have*

$$\tau_0 = \tau_2 + \sum_{d=1}^{L} \mathbb{E}\left[\tau_{1,d} | D = d\right] \mathbb{P}\left(D = d\right).$$

Proof. Clearly, $\tau_0 = \tau_1 + \tau_2$, where $\tau_1 = \mathbb{E}\left(Y(1,D) - Y(1,0)\right)$. And

$$\begin{aligned}
\mathbb{E}\left(Y(1,D) - Y(1,0)\right) &= \mathbb{E}\left[\sum_{d=0}^{D} \mathbf{1}_{\{D=d\}}(Y(1,d) - Y(1,0))\right] = \sum_{d=1}^{D} \mathbb{E}\left[\mathbf{1}_{\{D=d\}}\tau_{1,d}\right] \\
&= \sum_{l=1}^{D} \mathbb{E}\left[\tau_{1,d} | D = d\right] \mathbb{P}\left(D = d\right).
\end{aligned}$$

□

In other words, τ_0 the overall effect of Process 1 versus Process 0 is equal to τ_2, the intrinsic effect of Process 1 plus a weighted average of the intrinsic conditional effect of the policies.

This set up is similar to the framework of a randomized experiment with encouragement (Hirano et al. (2000)). Indeed, in designs with encouragement, individuals are encouraged to take a particular treatment but are ultimately free to choose whichever treatment they deem best.

Similarly in the design above, communities can choose any policy in a set $\{0,\ldots,D\}$. But there are some important differences here compared to designs with encouragement. First, because the chosen treatment is decided through a group decision-making process, the variable D is much more predictable compared to most designs with encouragement. The literature on group decision-making processes can be used to formulate interesting models describing D. Another difference is that contrary to what is typically done in the encouragement design literature, the *inclusion-exclusion* assumption is not acceptable here. In the framework above, the inclusion-exclusion assumption consists in assuming that $\tau_2 = 0$. As a matter of fact, the causal effect τ_2 is the prime effect of interest here as its measures the intrinsic effect of Process 1.

As in the encouragement design literature, the causal effect τ_0 can be seen as an Intent-To-Treat estimator which focuses on the causal effect of the assignment, rather than the causal effect of the treatment (policies). But the complication here is that in additional to the individual effect of each policy, τ_0 also contains the intrinsic effect of Process 1.

2.2 Statistical estimation

The causal effect τ_0 is estimable from the design. If the assignment to the two processes is randomized, then τ_0 can be estimated by comparing the average outcome over the communities under Process 1 and the communities under Process 0. But τ_2, the intrinsic effect of Process 1 cannot be estimated without further assumptions. For example, a simple comparison of the outcome of the communities under Process 1 that have selected Treatment 0 and the communities under Process 0 will not give τ_2 in general. Unless there is some strong evidence to believe that the choice of the policy is ignorable that is, does not depend on the outcome.

In order to separate τ_1 and τ_2, we assume that there exists some covariates $X = (X_1, \ldots, X_q)$ such that $Y(1,d)$ and D are conditionally independent given X. This is the strong ignorability assumption of Rubin & Rosenbaum (1983).

(A):

$$\mathbb{E}\left[Y(1,d)\mathbf{1}_{\{D=d\}}|X\right] = \mathbb{E}\left(Y(1,d)|X\right)\mathbb{P}\left(D=d|X\right), \quad d = 0,\ldots,L.$$

Then we define the propensity score function

$$\alpha_i(x) := \mathbb{P}\left(D = i|X = x\right).$$

Proposition 2.2. *Assume (A). Suppose that $\alpha_i(x) > 0$ almost surely. Then*

$$\frac{Y(1,d)\mathbf{1}_{\{D=d\}}}{\alpha_d(X)}$$

and $Y(1,d)$ have the same expectation.

Proof. The proof is a straightforward application of (A) and by conditioning on X

$$\begin{aligned}
\mathbb{E}\left[\frac{Y(1,d)\mathbf{1}_{\{D=d\}}}{\alpha_d(X)}\right] &= \mathbb{E}\left[\mathbb{E}\left(\frac{Y(1,d)\mathbf{1}_{\{D=d\}}}{\alpha_d(X)}|X\right)\right] = \mathbb{E}\left[\frac{1}{\alpha_d(X)}\mathbb{E}\left(Y(1,d)\mathbf{1}_{\{D=d\}}|X\right)\right], \\
&= \mathbb{E}\left[\mathbb{E}\left(Y(1,d)|X\right)\right] = \mathbb{E}(Y(1,d)).
\end{aligned}$$

□

We suppose that assumption (A) holds. Suppose now that we can collect data on N independent communities. Community k has variables

$$\mathcal{C}_k = \left(Z_k, D_k, Y_k(0), Y_k(1,0), \ldots, Y_k(1,L)\right).$$

$Z_k = 0$ if that community is assigned to Process 0 and $Z_k = 1$ otherwise. D_k denotes the policy chosen by the community and $Y_k(0)$ is the counterfactual outcome under Treatment 0 and $Y_k(1,d)$, the counterfactual under policy d under Treatment 1. We assume that the variables $\mathcal{C}_k$ are independent with the same distribution and for each k, the initial assignment Z_k is completely randomized. That is Z_k and $(D_k, Y_k(0), Y_k(1,0), \ldots, Y_k(1,L))$ are independent. For the k community, we observe Z_k, D_k and Y_k, where Y_k is defined as

$$Y_k = Y_k(0)\mathbf{1}_{\{Z_k=0\}} + \mathbf{1}_{\{Z_k=1\}}\sum_{d=0}^{L} Y_k(1,d)\mathbf{1}_{\{D_k=i\}}.$$

In other words, if $Z_k = 0$, we observe $Y_k = Y_k(0)$, if $Z_k = 1$ and $D_k = 0$, we observe $Y_k = Y_k(1,0)$ etc... Define the estimators

$$\hat{\tau}_N^{(0)} = \frac{\sum_{k=1}^N Y_k\mathbf{1}_{\{Z_k=1\}}}{\sum_{k=1}^N \mathbf{1}_{\{Z_k=1\}}} - \frac{\sum_{k=1}^N Y_k\mathbf{1}_{\{Z_k=0\}}}{\sum_{k=1}^N \mathbf{1}_{\{Z_k=0\}}},$$

$$\hat{\tau}_N^{(1)} = \frac{\sum_{k=1}^N Y_k\mathbf{1}_{\{Z_k=1\}}}{\sum_{k=1}^N \mathbf{1}_{\{Z_k=1\}}} - \frac{\sum_{k=1}^N \alpha_0^{-1}(X_k)Y_k\mathbf{1}_{\{Z_k=1\}}\mathbf{1}_{\{D_k=0\}}}{\sum_{k=1}^N \mathbf{1}_{\{Z_k=1\}}},$$

$$\hat{\tau}_N^{(2)} = \frac{\sum_{k=1}^N \alpha_0^{-1}(X_k)Y_k\mathbf{1}_{\{Z_k=1\}}\mathbf{1}_{\{D_k=0\}}}{\sum_{k=1}^N \mathbf{1}_{\{Z_k=1\}}} - \frac{\sum_{k=1}^N Y_k\mathbf{1}_{\{Z_k=0\}}}{\sum_{k=1}^N \mathbf{1}_{\{Z_k=0\}}}.$$

We make the convention that $0/0 = 0$. Given Proposition 2.2, it is easy to see from the expression why the estimates should be consistent. That is as N converges to infinity, $\mathbb{E}(\hat{\tau}_N^{(l)})$ should converge to τ_l.

Theorem 2.1. *Assume (A). Then $\mathbb{E}(\tau_N^{(l)})$ converges in probability to τ_l as $N \to \infty$. Moreover there exists $\sigma_l^2 \geq 0$ such that*

$$\sqrt{N}\left(\hat{\tau}_N^{(l)} - \tau_l\right) \xrightarrow{w} \mathcal{N}\left(0, \sigma_l^2\right), \quad \textit{as } N \to \infty. \tag{4}$$

Remark 2.1.

1. In other words, under (A), we can estimate and test the significance of each of the causal effects τ_l, provided N is large enough.

2. In practice $\alpha_0(x)$ is not known and needs to be estimated. We can do this using multinomial logit or probit regression. Knowledge of group decision theory can be useful to formulate such a model.

3. If we replace the function $\alpha_0(x)$ by an estimator in the estimators $\hat{\tau}_N^{(l)}$ above, what can we say about the asymptotic of the resulting estimators? It was shown by Hirano et al. (2003) that under some regularity conditions, the resulting estimator is still consistent, asymptotically normal.

2.3 Individual level model: partisanship effects

In most cases, this type of fields experiments have a limited number of treatment and control cases. Since the statistical analysis presented above relies on the number N of communities growing to infinity, it will not always be useful. With a limited number of cases, a more fruitful approach would be to work with individual data. Individual data also offers the possibility to investigate partisanship effects. Indeed the intrinsic effect of the process depends in general on how polarized the population is about the choice of the best policy.

To model these effects, we introduce the counterfactual variable $Y_i(1, d, 1)$ (resp. $Y_i(1, d, 0)$) as the outcome of individual i under Process 1 with group decision d and individual choice d (resp. different from d). In other words, if individual i is part of the majority that has selected d, we observe $Y_i(1, d, 1)$ otherwise, we observe $Y_i(1, d, 0)$. If we call M_i the indicator variable equal to 1 if i is part of the majority and 0 otherwise, we can define

$$Y_i(1, d) = Y_i(1, d, 1)\mathbf{1}_{\{M_i=1\}} + Y_i(1, d, 0)\mathbf{1}_{\{M_i=0\}}.$$

We have

$$\tau_2 = \mathbb{E}\left(Y_i(1,0,0) - Y_i(0)|M_i = 1\right)\mathbb{P}\left(M_i = 1\right) + \mathbb{E}\left(Y_i(1,0,1) - Y_i(0)|M_i = 0\right)\mathbb{P}\left(M_i = 0\right).$$

In other words, the intrinsic effect of the process can be written as the average of the average intrinsic effect among those that have voted from policy 0 and those that have voted against it.

3 Town meeting campaign experiment in Benin

The town meeting campaign experiment investigates the effect of public debates around specific and informed policy platforms on turnout and voting, in the context of 2006 presidential elections in Benin. [4]The experiment has two main stages. The first stage involves political parties and is led by policy experts. The second stage draws on the outcome of the first stage, and consists of town meetings with voters, led by party activists. Thus, The treatment is *not* a pre-designed, pre-crafted platform or a vignette that would be read to voters. Instead, it is a process, a *mechanism* for generating political platforms or campaign messages.

The experimental process started with a policy conference that took place on December 22, 2005, entitled "Elections 2006: What policy alternatives?". There were about forty participants and four panels (Education, Public Health, Governance, and Urban Planning). Four policy experts wrote reports describing government performance in those four areas and outlined recommendations to candidates and parties, based on academic research and best practices [5] After the conference several political parties and candidates volunteered to experiment with the proposed campaign strategies in 14 randomly selected villages.

Once the assignment of electoral districts to treatment and control groups was completed, teams of campaign workers were instructed with specific policy responses to voters' concerns. . They were also given specific instructions on how to run the town meetings: First, they introduced themselves and the candidate they were representing. Next, they gave a fifteen minute speech on the key problems facing the country and on the specific solutions suggested by the candidate. The speech triggered an open debate in which the issues raised were contextualized, and the proposals made were amended by the participants. The meeting lasted between ninety minutes and two hours.

[4] The first part of the section draws mostly from Wantchekon (2008)

[5] The four experts were Professor Leonard Fourn who teaches Public Health at the University of Abomey Calavi, Dr. Hamissou Oumarou, an Education Expert from Niger, Dr Mouftaou Laleye, who taught Public administration at the University of IFE in Nigeria, and Mr Todjinou Jean Bosco, an architect and Urban Planning specialist.

While villages in treatment groups received and deliberated in town meetings over informed and broad-based policy proposals; villages in the control groups received clientelist campaign promises, in the context of rallies.

On election day, data were collected at the relevant sites on turnout in treatment and control precincts with the help of representatives of the National Electoral Commission. A week after the elections, a representative sample in each group was surveyed on demographic variables (age, gender, marital status and ethnic affiliation), socioeconomic variables (educational attainment, economic activities, and assets) and political variables (preferences over candidates and voting behavior).

3.0.1 Main results

For our purpose, the two main outcomes of interest are voter information and voting behavior. In the post-election survey, voters were asked the following three questions: (1) Did the campaign give you information about the quality of the candidates? (2) Did the campaign give you information about government and how it functions? (3) Did the campaign give you information about the problems facing the country?

The question that best captures the concept of voter information is the one on the problems facing the country and, to a lesser degree, the one on the quality of the candidates. Information on governments is a measure of the level of civic education rather than a measure of voter information. Thus, we will focus our attention on (1) and (3)

Voter Information Tables 1A and 1B from Wantchekon (2008) present the results on policy and candidates information. In all specifications except one, the treatment has a positive and significant effect on policy information. The results are significant at the 99% level without clustering and the 90% level with clustering. As for information about the candidates, the treatment has a positive effect in all specifications. The results are significant at the 99% level without clustering and the 95% level with clustering. Education and gender are highly correlated with voter information. More specifically, male voters are more likely to find the campaign informative with regards to policies and candidates. Ethnic ties are a good predictor of voter information about candidates, but not about policies, and media access has no significant effect.

Insert Tables 1A and 1B here

Now, let us turn to the treatment effect on voting.

Voting behavior Table 2A uses data collected from the electoral commission on the outcome of the election in treatment and control villages. Overall, the experimental candidates garnered 66.7% of the vote in the treatment villages, compared with 60.7% in the control villages. In one commune (Kandi) the results were approximately the same for the experimental and control villages. In four out of seven cases, the experimental candidate gained more votes in the treatment villages, with the treatment effect being particularly strong in Gadome I and Yaoui.

However, in two districts out of seven, the experimental candidates fared better in the control villages. For instance, in Kouande, the experimental candidate gained a slightly higher percentage of votes in the control group than in the treatment. This may be explained by an unexpected rally by the candidate participating in the experiment in that district, Yayi Boni, just two days before the election. There were no such rallies in any other district participating in the experiment.

Insert Table 2A here

There are three districts that took part in the experiment that are missing from Table 4A because we could not get an accurate vote count from these districts on election day (Abomey Calavi, So Ava and Zagnanado). In these districts, the participating political party was Renaissance du Benin and the experimental candidate was Lehadi Soglo. In sharp contrast with the other experimental candidates, Soglo was the underdog in each district. Table 4B presents the vote shares of the candidate using the post-election surveys. The results are strikingly similar to the one described in Table 4A. In two districts (So Ava and Abomey Calavi) the treatment effect is positive and in one district (Zangnanado) the effect is negative. Thus there is a strong indication that the average treatment effect would have remained positive had we included the three missing districts.

Insert Table 2B here

Wantchekon (2008) also uses a probit model to test the effect of the treatment on voting.

$$
\begin{aligned}
P\left(Y_{ij}=1|z_{ij},T_i\right) &= P\left(z_{ij}a+T_i\beta+x_{ij}T_i\gamma+u_{ij}>0\right) \\
& \quad u_i \overset{id}{\sim} N(0,\Omega_i)
\end{aligned}
$$

But here, Y_{ij} is a categorical variable that takes the value of one if individual j in village i votes for the "experimental"candidate in the 2006 election and zero otherwise, z_{ij} is the vector of individual characteristics for individual j in village i, and T_i is the categorical variable for treatment in village

i. The vector z_{ij} includes variables such as age, gender, level of educational attainment, ethnic ties with the candidate, and media access. Income level was measured by using an index of housing quality, constructed from factor analysis of five independent variables (roofing, ground, number of rooms, etc.). The key independent variable is T_i, the treatment, which takes the value of one if the respondent was in the treatment group and zero if the respondent was in the control group.

In each specification, we present the results without any controls, then we control for the two covariates that are not balanced between treatment and control groups (i.e. education and media access). Finally, we control for all potentially relevant covariates.

Table2C indicates that the treatment has no effect on voting behavior, which is a bit surprising given the results described in Table 2A and 2B. This is probably due to the fact that the post-election survey data was collected a week after the election and two days after the results were announced. Yayi Boni, the main experimental candidate, won the first round of the election by ten points, and it is likely that respondents in areas where he did less well might have exaggerated their electoral support for him after learning the results. For instance, in the districts where we ran the experiment, Yayi's vote share is 31% higher in the post-election survey than in the election-day vote count. Thus, if he were to do better in treatment areas than in control areas on election day, this margin would be much narrower after the results were announced. It is therefore safe to conclude that the results in Table 4C underestimated the effect of the treatment on voting behavior.

Insert Table 2C

However, compared with the results of the 2001 experiment, where programmatic platforms had a negative and significant effect on electoral support, we can claim the town meeting and the specificity of promises helped at least to close the gap between programmatic platforms and clientelist platforms.

3.0.2 Strategies to separate out the effect of town meetings from the effect of information

The task here is to try and separate the causal effect of the specific platform from the causal town meetings. Given the design of the experiment, this is not rigorously possible. There is no treatment group where specific programmatic platforms were delivered at rallies. But we can make two assumptions that would help disentangle the two effects. The first is that both control and treatment platforms are "specific". The only difference being that the treatment platform focuses on the problems facing the country and control platforms focus on the problems of the village. In addition, based on the results of the 2001 experiments, we assume that clientelist specific platforms would gather at least much support as programme-specific platforms platforms. This means that clientelist specific platforms would have done at least as well in town meetings as programme-specific platforms and programme-specific platforms would have done no better than a clientelist specific platform. Under this assumption, any difference in terms of voting between the treatment and control groups, would due to the town meeting, especially the difference between those in both groups who found the campaign informative about policies and candidates. Therefore we can consider the result in Table 3C as a lower bound on effect of town meeting. In other words, even if there is no significant difference between treatment and control in terms of voting behavior, we argue that the effect would have been positive if we have ran programme specific platforms in control areas.

In order to derive the information effect, we can focus only on the treated villages and assume that assigning a village to treatment implies that all inhabitants of that village have been exposed to town meeting. We also assume that everybody in treated village has been encouraged to take the programme-specific platform, i.e. content of the town meetings. We then have the structure of a design with encouragement. Let $D_i = 1$ if i complies with the encouragement and $D_i = 0$ otherwise. Since we don't observe D, we take D as the variable 'Did the campaign give you information about the problems facing the country'. We can then estimate the propensity score $p(x) = \mathbb{P}(D = 1|X = x)$ using a logistic regression. Using this propensity score, we can estimate the causal effect of the 'content of the town meeting' as

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{Y_i D_i}{\hat{p}(X_i)} - \frac{Y_i (1 - D_i)}{1 - \hat{p}(X_i)} \right).$$

We can then take $\hat{\tau}$ as the causal effect of the content of the town meetings.

Table 3, indicates that in the treatment group, information has a positive effect on voting for the candidate. We result are confirmed when we estimate $\hat{\tau}$ (not shown)

4 The Democracy Experiment in Indonesia

To be completed

5 Guidelines for experiments involving institutions

To be completed

6 Conclusion

We propose a framework for estimating the intrinsic impact of a decision-making process(or institution) in experiments where such a process is randomly assigned to groups of individuals who then decide which treatment to receive. In our framework, a randomized evaluation of institutions has the structure of group-based encouragement design with multiple choice over treatments or policies. The main challenge in such experiments is to separate the institutional effect from the policy effect.

Our empirical strategy consists first, of estimating the propensity to adopt a policy among individuals in the treatment group. Then, assuming that policy selection is conditional only on observed covariates, we can compute the policy effect. Finally, we can derive the institutional effect by subtracting the estimated policy effect from the "total" treatment effect,i.e. the difference in means between treatment and control group observations.

Our results could help improve our understanding of how results from policy experiments would change when they are brought to scale, when institutional constraints are integrated into the analysis. In addition, with the ever growing interest of social scientists in the effect of institutions, our paper contributes to this effort by proposing an experimental strategy for estimating the direct effect of institutions on behavior.

References

[1] Adams, William C. and Smith J. Dennis. 1980. Effects of Telephone Canvassing on Turnout and Preferences: A Field Experiment. *The Public Opinion Quarterly,* Vol. 44, No. 3 pp. 389-395

[2] Angrist, Joshua, Guido Imbens and Donald Rubin. 1996. Identification of Causal effects Using Instrumental Variables. *Journal of Econometrics*, Vol. 71,No. 1-2, 145-160

[3] Banerjee Abhijit and Esther Duflo (2008),"The Experimental Approach to Development Economics," Forthcoming *Annual Review of Economics*, (also see CEPR working paper No. DP7037, NBER working paper No. 14467)

[4] Duflo, Esther. 2006. Field Experiments in Development Economics. Working paper, MIT

[5] Fisher, R. (1935). The design of experiments. Boyd, London

[6] Gerber, Alan, and Donald P. Green. 2000. "The Effects of Canvassing, Phone Calls, and Direct Mail on Voter Turnout: A Field Experiment" *American Political Science Review*,Vol. 94, No. 3; pp. 653-663

[7] Gerber Alan and Donald Green. 2007. "Field and Natural Experiments" Forthcoming. *Handbook of Political Methodology* (Chapter 38)

[8] Glewwe, Paul, Michael Kremer, Sylvie, and E. Zitzewitz (2004). "Retrospective vs. ProspectiveAnalyses of School Inputs: The Case of Flip Charts in Kenya," *Journal of Development Economics*. Volume 74(1), pp. 251-268.

[9] Gosnell, Harold F. 1927. *Getting-Out-the-Vote: An Experiment in the Stimulation of Voting.* Chicago: University of Chicago Press.

[10] Harrison Glenn, Morten Lau and Elisabet Rutström 2009 "Risk Attitudes, Randomization to Treatment, and Self-Selection Into Experiments," *Journal of Economic Behavior and Organization,* forthcoming.

[11] Hirano, K., Imbens, G. and Ridder, G. (2003). Efficient estimation of average treatment effect using the estimated propensity score. *Econometrica* 71 1161-1189[

[12] Hirano Keisuke., Guido. Imbens, Donald. Rubin, and Xiao-Hua. Zhou. 2000. "Assessing the Effect of an Influenza Vaccine in an Encouragement Design with Covariates," *Biostatistics* 1, 69-88.

[13] Holland, P. (1986). Statistics of causal inference. Journal of American Statistical Association 81 945-970

[14] Miguel, Edward and Michael Kremer (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica*, Volume72 (1), pp. 159-217

[15] Miller, Roy E., David A. Bositis, and Denise L. Baer. 1981. "Stim- ulating Voter Turnout in a Primary: Field Experiment with a Precinct Committeeman." *International Political Science Review* 2 (4): 445-60.

[16] Imbens Guido and Donald Rubin. 1997. Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance. *Annals of Statistics*, Vol. 25, No. 1, 305–327

[17] Neyman J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9 (with discussion) translated in Statistical Sciences Vol 5, No 4 465-480

[18] Olken Benjamin, 2008. Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia NBER Working Paper #14123.

[19] Rubin, D. (1974). Estimating causal effects of treatments in randomized and non randomized studies. *Journal of Educational Psychology* 66 688-701

[20] Rosenbaum, P. R. and Rubin, D.(1983). The central role propensity score in observational studies for causal effects. *Biometrika* 76 41-55

[21] Wantchekon, Leonard. 2008. Clientelism and voting and behavior: Evidence from a field experiment in Benin *World Politics* 55 399-422

[22] Wantchekon, Leonard. 2008. Expert Information, Public Deliberation, and Electoral Support for Good Governance: Experimental Evidence from Benin. Working Paper, New York University

Table 1A: Information - Candidates

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	.169***	.169**	.167***	.167**	.156***	.156**
	(.055)	(.066)	(.056)	(.072)	(.058)	(.061)
Education			.314***	.314***	.198***	.198***
			(.59)	(.075)	(.064)	(.076)
Gender (male=1)					.351***	.351***
					(.061)	(.055)
Age					-.001	-.001
					(.002)	(.002)
Ethnic Ties					.487***	.487*
					(.086)	(.288)
Media			-.281***	-.281*	-.245***	-.245
			(.054)	(.166)	(.061)	(.168)
Discussion					-.211***	-.211**
					(.037)	(.091)
Observations	2073	2073	2073	2073	2052	2052
Pseudo R^2	.015	.015	.034	.034	.079	.079
Clustered Standard Errors	No	Yes	No	Yes	No	Yes

Note: The estimation method is probit. Standard errors in parentheses. Clustering is at the Commune level. All models include candidate fixed effects. *significant at 10%; **significant at 5%; ***significant

at 1%

Table 1B: Information - Problems Facing Country

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	.153***	.153*	.143**	.143	.177***	.177*
	(.058)	(.091)	(.058)	(.094)	(.060)	(.104)
Education			.426***	.426***	.339***	.339***
			(.061)	(.064)	(.065)	(.071)
Gender (male=1)					.236***	.236***
					(.063)	(.063)
Age					.002	.002
					(.002)	(.003)
Ethnic Ties					-.016	-.016
					(.092)	(.193)
Media			-.151***	-.151	-.014	-.014
			(.057)	(.153)	(.064)	(.116)
Discussion					-.288***	-.288**
					(.039)	(.121)
Observations	2073	2073	2073	2073	2052	2052
Pseudo R^2	.046	.046	.066	.066	.099	.099
Clustered Standard Errors	No	Yes	No	Yes	No	Yes

Table 2A: Vote Shares of Experimental Candidates (official results)

Commune	Village	Party	Status	Vote shares.	Vote Total
Kandi	Thya	UDS	T	71.5	601
			C	72.8	29,524
Bembereke	Mani	UDS	T	64.3	193
			C	73.3	24,007
Ouesse	Yaoui	CAP	T	80.4	1,495
			C	62.7	24,186
Save	Okounfo	CAP	T	72.0	713
			C	61.6	20,314
Come	Gadome I	IPD	T	54.3	578
			C	32.3	8,500
Dangbo	Mitro	PRD	T	59.4	413
			C	54.1	2509
Kouande	Orou-Kayo	IPD	T	60.7	482
			C	68.3	17160
Tanguieta	Taicou	IPD	T	25.98	1216
			C	22.42	1320

Note: T means Treatment and C means Control

Table 4 C: Vote for Experimental Candidate

	(1)	(2)	(3)	(4)
Treatment	-.025	-.019	-.050	-.181
	(.286)	(.284)	(.278)	(.205)
Education		-.247**	-.227**	-.253
		(.119)	(.107)	(.159)
Media		.059	.011	.316
		(.218)	(.198)	(.255)
Gender (male=1)			-.095	-.059
			(.061)	(.107)
Ethnic Ties			.742***	.639**
			(.277)	(.327)
Treatment* Media				-.578*
				(.351)
Treatment*Gender				-.081
				(.137)
Treatment*Ethnic Ties				.234
				(.476)
Treatment*Education				.043
				(.164)
Observations	2058	2058	2058	2058
Pseudo R^2	.374	.379	.391	.399

Note: The estimation method is probit. Standard errors in parentheses, clustered at the Commune level. All models include candidate fixed effects. *significant at 10%; **significant at 5%; ***significant at 1%

Table 3: The Effect of Information in Treatment Areas

	(1)	(2)	(3)
Informative Campaign	.134	.202**	.097***
	(.091)	(.089)	(.022)
Gender (male=1)		-.209***	-.189***
		(.064)	(.073)
Age		.007	.005
		(.006)	(.005)
Education		-.179	-.128
		(.116)	(.126)
Ethnic Ties			1.186***
			(.316)
Media			-.194
			(.209)
Housing			.057
			(.141)
Constant	.518***	.345	.676
	(.059)	(.296)	(.441)
Pseudo R^2	.349	.354	.382
Observations	1026	1019	993

Note: DV is vote for experimental candidate. The estimation method is probit. **significant at 5%; ***significant at 1 %; Clustered standard errors in parentheses. All models include candidate fixed effects.