

# **Design and Analysis of Experiments in Multi-level Populations**

Betsy Sinclair (Chicago)

May 18, 2009

The author would like to thank Donald Green, Jamie Druckman, Thomas Leeper, Jon Rogowski, John Balz, Alex Bass, Jaira Harrington, and participants of the West Coast Experimental Political Science Conference for their helpful comments in improving this manuscript. Comments should be directed to [betsy@uchicago.edu](mailto:betsy@uchicago.edu).

# 1 Introduction: Statement of the Problem

Randomized experiments are seen as among the most rigorous methodology for testing causal explanations for phenomena in the social sciences and are experiencing a resurgence in political science. The classic experimental design randomly assigns the population of interest into two groups, treatment and control. Ex-ante these two groups should have identical distributions in terms of their observed and unobserved characteristics. Treatment is administered based upon assignment, and by the assumptions of the Rubin Causal Model the effect of the treatment is calculated as the difference between the average outcome in the group assigned to treatment and the average outcome in the group assigned to control.

Randomized experiments are often conducted within a multi-level setting. These settings can be defined both at the level at which the randomization occurs as well as at the level at which the treatment is both directly and indirectly administered. These indirect effects most often occur as a result of social transmission of the treatment, which is particularly likely when the treatment consists of information. This essay will explore the implications of these multi-level settings to highlight the importance of careful experimental design with respect to random assignment of the population and the implementation of the treatment. There are potential problems with analysis in this context and this essay suggests strategies to accommodate multi-level settings. These problems are most likely to occur in field settings where control is lacking although it can sometimes occur in other settings as well.

There are two common environments where the standard experiment fails to adjust for a multi-level structure and results in problematic estimates and both occur where the assignment to treatment is at the individual level but the administration of treatment is not. Both of these instances create problems for analysis. The first problem relates to the selection of groups to receive treatment. Suppose the random assignment is conducted at the individual level, but the implementation of the treatment is directly at the group level and suppose that the selection of which groups to treat is not random but instead

is selected by an organization where the selection is correlated with individual voting probabilities. This produces bias in the inferences that result from this setup.<sup>1</sup> An example of this type of experiment would be one in which the randomization assigns individuals to treatment and control but administers treatment to particular ZIP codes. Inferences in this context are problematic because of the selection of particular ZIP codes.

The second problem is in terms of social interactions – we could have possible spillover effects from one individual to another within the same household for example, or furthermore from one household to another within the same group. That is, the random assignment is conducted at the individual level, but the implementation of the treatment is indirectly at the group level. Again, an example of this type of experiment would be one in which the randomization assigns individuals to treatment and control but administers treatment to households. Inferences at the individual-level are then problematic because the treatment can be socially transmitted within the household. Social science randomized experiments often rely upon treatments which can be socially transmitted. Social transmission has the potential to result in violation of one of the fundamental assumptions in the analysis of these experiments, the Stable Unit Treatment Value Assumption (SUTVA). SUTVA states that there is no interference between units, such that the assignment of an individual to the treatment group should have no effect on outcomes for other individuals. Many randomized field experiments take place in situations where interference between individuals is likely, such as within social settings where social interaction is expected. If ignored, SUTVA violations have the possibility of adding bias to estimated treatment effects, and it is possible that these biases can go in either a positive or negative direction.

Multi-level randomized experiments rely upon existing social structures, which have the potential to provide solutions to the problems that arise from social transmission. A multi-level randomized field experiment design allows both for the assurance of no SUTVA violations as well as an explicit test for the degree to which social interactions take

---

<sup>1</sup>As oftentimes not all individuals who are assigned to receive the treatment will actually be treated, there is also a loss of efficiency if the contact rates differ greatly across groups.

place. Making SUTVA an object of study instead of an assumption has the benefit of providing new insights about interpersonal influence. Multi-level experiments provide an opportunity to better understand social transmission of politics. While theories abound about the structure of social environments, little empirical evidence exists to explicitly validate that these structures influence an individual's politics. Multi-level settings occur when individuals are assigned to treatment but will communicate to each other within different social levels, such as within households or within precincts. Multi-level settings require specific experimental designs and analyses but allow us to estimate the effects of interpersonal interactions. This essay describes both the advantages of a multi-level randomized field experiment and provides recommendations for the implementation and analysis of such experiments consistent with the current best practices in the literature.

## **2 Spillovers, Models of Diffusion, and the Reflection Problem**

Multi-level contexts highlight the role that social interactions may play in political behavior. Turnout decisions may diffuse through a population in much the same way that disease or trends are also transmitted across individuals who are socially connected. Diffusion processes have been clearly seen in marketing effects (Coleman, Katz and Menzel 1966, Bhatia, Manchanda and Nair 2008) and are likely to also be present in political behavior. Formal models of spillover effects in political behavior suggest that the ways in which individuals are socially connected are highly likely to determine their beliefs about candidates (Sinclair and Rogers 2009, DeMarzo, Vayanos, Zwiebel 2003).

These models are difficult to estimate using observational data due to the reflection problem (Manski 1993). This is an identification problem, where it is difficult to separate the selection into group membership from the effect of the group itself. Thus observationally individuals appear to be correlated in terms of behavior with their peers when the cause of this correlation could be either a factor of the impact of group members on the individual, a common factor which affects all members of the group, or additionally

affects characteristics which the members of the group are likely to share and which have driven their selection into the group initially. Standard regression models set up to provide critical tests of these formal models are exposed to the reflection problem, as it is impossible to determine whether or not the individual is having an effect on the group or vice versa, for example. Alternative estimation strategies do exist, which include the use of instrumental variables and the role of time (Brock and Durlauf 1999, Conley and Udry 2009). Yet observational strategies are not generally sufficient to identify causal effects of social interaction.

By using randomized experiments, we are able to gain some leverage on the reflection problem by providing a stimulus to one set of individuals and then observing the behavior of the group. In particular, by looking for empirical evidence of spillovers in this setting, we are then more able to directly test to what extent models of diffusion are applicable to political behavior. By using a multi-level randomized experiment, we are able to directly evaluate both the effect of the treatment and peer-effects. This occurs via the identification of the appropriate multi-level context and is described in the next section.

With limited exceptions, random assignment is typically done at either the individual or population level in randomized field experiments. There are many examples where randomization occurs at the population level, ranging from assignment by neighborhood, congressional district, or precinct, often because it is difficult to administer treatment by individual (Arceneaux 2005, Imai King and Nall 2009, King et al 2007). For example, in experiments on the effects of campaign advertising, randomization often occurs at the level of the media market (Green and Vavreck 2008, Panagopoulos and Green 2006). Some experiments have failed at the individual level – for example, in experiments on policy such as antipoverty efforts (Adato, Coady and Ruel 2000) and reductions in class size (Krueger 1999) – because the subjects were able to change their own assignment category within a particular group, thus some types of experiments must be conducted at the population level. If randomization occurs at the population level, then without additional assumptions the administration of treatment and inferences about treatment efficacy will also be

done at the population level. One difficulty in evaluating the treatments in these contexts is that individuals may have communicated to each other about the treatment, thus the observed treatment effect may result from an interaction between the direct and indirect administration of treatment. This has implications for both the external and internal validity of the experiment, as it is not possible to separate the direct and indirect treatment effects.

Randomization occurs at the level of the individual as well, for example, in the bulk of experiments on voter mobilization tactics (See Gerber and Green 2008 for a review of the literature). The majority of this essay will address issues involving the estimation of direct and indirect treatment effects when inferences are drawn about the behavior of the individual. It is crucial to account for the level at which randomization occurs. Inferences are generally drawn about the behavior of the individual and it is possible to create inconsistencies between the assignment of individuals to treatment and the implementation of the treatment.

Implementation of the treatment may occur indirectly as the result of social interaction. Many randomized experiments administer treatments which could be socially transmitted, such as political information, a heightened sense of political interest, or increased social trust. Empirical work on social networks has suggested that many of these treatments can be socially transmitted (Nickerson 2008, Cacioppo, Fowler and Christakis 2009, Fowler and Christakis 2008). Experiments where the treatment is subject to social transmission are implicitly conducted within a multi-level setting.

### **3 The Stable Unit Treatment Value Assumption**

Within a social setting it is possible that spillovers from individuals assigned to treatment to individuals assigned to control could create biases in the estimation of treatment effects, or vice versa. Here we consider the standard setup for a political interest experiment and examine ways in which these biases can be eliminated. Here our narrative, consistent with work by McConnell, Sinclair and Green (2009), considers an experimental popula-

tion where individuals are members of a three-level multi-level setting. Individuals are residents in a household and each household resides within a social group. The collection of groups form the population. This example could generalize to any number of levels or different settings, so long as each sub-level is fully contained within the previous level.

Our primary concern with the assignment and implementation of randomized field experiments within a multi-level population is violations of the stable unit treatment value assumption (SUTVA, as labeled in Rubin 1980). Units here are defined as the unit which is being evaluated (for example, if the treatment is being evaluated at the individual level, then the individual is the unit, whereas if the treatment was being evaluated at the group level, then the group is the unit).

SUTVA states that the potential outcomes for any unit do not vary with the treatments assigned to any other units, and there are no different versions of the treatment (Rubin 1990). The assumption of SUTVA is key to how we draw causal inferences about the efficacy of the treatment. The first part of SUTVA assumes that there is no interference between units; that is, it assumes that there are no spillover effects. The second part of SUTVA assumes that the treatment is the same for each unit. Says Rubin, "SUTVA is violated when, for example there exist unrepresented versions of treatments or inference between units" (Rubin 1986, p. 961). In this essay, we will focus exclusively on the problem of inference between units as a result of treatment, specifically treatment spillovers.

It is possible that the units do interfere with each other in the course of receiving treatment. For example, suppose that some individuals are contacted and given additional political information – it seems likely that they might then discuss this information with the other individuals they know in their household or in their neighborhood. In this case there would be interference between units. In this essay, we will investigate the extent to which we can measure potential spillovers and also design experiments in order to be able to correct for their potential effects. We contend that spillover effects could exist within households or within groups.

We now look at an example where, in the presence of spillovers, it would not be pos-

sible to ascertain the exact treatment effect if the randomization is conducted at the individual level and there is communication between individuals within the experimental population. We explore spillovers both within their households and within their groups. That is, suppose we have a violation of SUTVA. We consider the violation in terms of the intent-to-treat effect (ITT).

### 3.1 Intent-to-Treat Effect

In our population of  $n$  individuals, let each individual  $i$  be randomly assigned to treatment  $t = 1$  or control  $t = 0$ . We investigate the outcome for each individual  $Y_i$ . We want to know what the difference is between treatment and control – the treatment effect – and ideally we would like to calculate  $Y_{i,t=1} - Y_{i,t=0}$ , yet it is not possible to observe both states of the world at the individual level. However, due to the random assignment the group of individuals who receive treatment are ex-ante identical in terms of their characteristics to those who receive control, and we are able to look at the difference in terms of expected outcomes. We define the intent-to-treat effect as  $ITT = E(Y_i|t = 1) - E(Y_i|t = 0)$ . The assumption of SUTVA allows us to consider only the assignment of individuals. However, when individuals are communicating, we must then revise our statements to include the assignments for the other individuals in our sample as well. That is, we must now revise our statement to describe all  $n$  individual assignments.

To consider a case where individuals are communicating within their households, suppose we consider a case where all individuals live in two-person households and the second person in the household is identified as  $j$ . When measuring the expected outcome we have to consider the assignment to the second person in the household as well,  $E(Y_i|t_i, t_j)$ . Thus in order to estimate the treatment effect, we need a particular group of individuals who have been assigned to treatment where the other individuals in their household have been assigned to control, so that  $ITT = E(Y_i|t_i = 1, t_j = 0) - E(Y_i|t_i = 0, t_j = 0)$ . Yet the standard inference would not have been written to incorporate the second person, so that there is a chance that the inferences could have been made based upon a situation where

the second person was in fact assigned to treatment as well, resulting in a biased effect due to communication between individuals. That is, suppose there are four individuals where three are assigned to treatment and one is assigned to control, but that we do not draw inferences based upon a multi-level structure. Then we could misestimate the treatment effect as  $ITT = 1/6*(Y_1|t_1 = 1, t_2 = 1) + 1/6*(Y_2|t_1 = 1, t_2 = 1) + 1/6*(Y_4|t_4 = 1, t_3 = 0) - 1/2 * (Y_3|t_3 = 0, t_4 = 1)$ . For individuals 1 and 2, they may be more likely to change their behavior because they both receive treatment, so this suggests that in fact we could overestimate the treatment effect, yet individual 3 may also be more likely to change behavior since she shares a household with someone who was also in the treatment group, so this suggests that we could in fact underestimate the treatment effect.

Extending this example to groups, we would then have an even more complicated problem where there could be communication between many households within a group. The true  $ITT$  would then need to be written based upon all the instances of communication.

The consequences of these spillovers are such that it is possible that in the presence of communication, the estimated treatment effect can be either an overestimate or an underestimate of the true effect. The direction of the bias will depend on the ways in which communication occurs and the effect of communication on an individual's decision – it may be the case that additional communications between treated individuals, for example, will heighten the probability that they change their behavior. Violations of SUTVA may produce biased estimates in light of communication about treatment. It is not possible to know whether or not the direction of the bias will be positive or negative prior to conducting the experiment.

## 4 Identifying a Multi-level Context

We identify two types of multi-level contexts. First, multi-level contexts are likely to exist where the randomization is conducted at a different level from the level at which treatment is administered. This type of multi-level context has the potential to generate

correlations within groups about the treatment. Experiments like these occur most often when the experimenter is relying upon the multi-level structure in order to implement the treatment. Examples include voter mobilization experiments conducted by precinct or household. The design of these experiments must account for this structure. If it is the case that there is likely to be any failure-to-treat – that is, if it is possible that not all the groups are going to actually receive the treatment – then it is essential that the order of the attempts to contact each group be randomized.<sup>2</sup> This randomization both allows for valid causal inferences and makes it impossible for the selection of particular groups to undermine the randomization. So long as all units will be treated, then the key in drawing inferences in these cases is that if treatment is administered at a different level than that of the randomization, the inferences must adjust for this correlation. This adjustment is discussed in the next section.

The second area where multi-level contests are likely to exist are where the treatment consists of something which can be communicated across social ties, such as information. This type of randomized experiment need only be considered where individuals in the population of study are members of the same social structure. That is, instances, where for example, it is possible that an individual assigned to control and an individual assigned to treatment could be residents in the same household. This type of multi-level context requires a very specific design as the treatment has the potential to be indirectly administered at the group level.<sup>3</sup> This case has the potential to be extremely problematic for drawing valid causal inferences without additional adjustment. This case has the potential to violate SUTVA.

Empirically scholars have observed social spillover which could generate SUTVA violations in the classic experimental framework. For an example of within-household interference, Nickerson (2008) finds higher levels of turnout in two-person households when

---

<sup>2</sup>Failure-to-treat problems present problems for analysis, some of which can be mitigated via randomization inference (Hansen and Bowers 2008).

<sup>3</sup>If we conduct both our randomization at the group level and our analysis on the group level, then this case requires no additional shifts in experimental design and in fact is eligible for the block-group randomization (King et al 2007).

one of the individuals is contacted to get out to vote via door-to-door canvassing in a voter mobilization experiment. In this instance, there is interference across within the households. Nickerson finds that 60% of the propensity to vote can be passed onto the other member of the household – a precise measurement of treatment spillover. Green, Gerber and Nickerson (2003) find within household spillover effects from door-to-door canvassing: an increase of 5.7 percentage points for other household members among households of younger voters. In one of the earliest mobilization experiments, the Erie County study reported that while 8% of Elmira residents had been contacted, turnout increased by 10%, suggesting that mobilization contact was socially transmitted (Berelson, Lazarsfeld and McPhee 1954). Other scholars have examined spillover effects in contexts unrelated to political behavior (Duflo, Kremer and Robinson 2006, Munchi 2004, Besley and Case 1994, Miguel and Kremer 2004). Each of these instances would generate a SUTVA violation.

In order to correctly identify instances where a multi-level structure exists to correct the experimental design to adjust for potential SUTVA violations, it is necessary to have additional information about the population of interest. There are several possibilities for identifying multi-level social structures, whose pragmatism is highly dependent upon the type of experiment being conducted. The first is to explicitly observe the level at which social interactions occur. For example, for researchers conducting experiments where they have clear and explicit knowledge of an individual's full network (for example, if the experiment was conducted via the social-networking website Facebook), then it is possible to explicitly conduct randomizations across separate components of personal networks so as to insure against spillover. However, most experimental frameworks do not allow for this type of explicit specification of the full network structure. An alternative method for observation is for researchers to rely upon survey results where individuals self-identify their social ties. Randomization can then occur within an individual's self-identified social relationships. Survey data which solicits an individual's discussion partners – people with whom they are likely to communicate with about the treatment and thus where spillovers are likely to occur – has demonstrated that many of these dis-

cussion partners are geographically proximate (McClurg 2006, Huckfeldt, Johnson and Sprague 2004). Thus, the final method for observation of an individual's social structure is geography. Relying upon geography requires no additional acquisition of data and also allows the researchers to investigate to what extent there is spillover within an individual's physical context. The choice of each of these methods – explicit observation, survey, and geography – should in large part be based upon the type of treatment administered.

## 5 Designing a Multi-level Experiment

The problem generated by communication of the treatment with participants in the experiment enables us to generate an alternative experimental design which relies upon our knowledge of an individual's social structure. A multi-level experimental design is the best approach for establishing causal inferences where experiments are conducted within an individual's social structure. Our proposed solution is to introduce additional randomization, consistent with recommendations by McConnell, Sinclair and Green (2009). With these additional randomizations, we are able to establish via the Rubin potential outcomes framework a treatment framework which can identify both the spillovers and the true treatment effect. These additional randomizations are key to designing a multi-level experiment.

The first component of such a design is to identify all levels at which individuals will indirectly interact. For purpose of example, we consider a population where individuals are likely to interact with each other on two levels, household and group. We require that these groups must be subsets of each other, so that each individual in the population is part of exactly one group and one household. The social structure of our population can be seen in Figure 1. Key to this analysis is to incorporate the full set of social structures, ranging from those where the most intimate interactions are likely to occur (in this case, households) up to those where the most casual interactions are likely to occur. In our example we model this as the group, but these group level interactions could truly be at the neighborhood level, the school-district level, or the city level. There must ex-

ist a group level at which individuals will not interact but levels where the experiment will take place and furthermore each level must be distinct – that is, an individual cannot belong to multiple groups. One way to ensure this is possible is to carefully select the experimental population; if an individual shares a household with individuals who are not part of the experiment, then it is not necessary to account for their presence in the estimation of the treatment effect, so individuals should be eligible to participate in the experimental population if their group memberships can be summarized by the appropriate randomization-levels. It is also possible to increase the number of randomizations to include, as an additional category, those individuals who belong to multiple groups, although this significantly increases the number of individuals necessary to incorporate into the experiment. Identifying these social structures and identifying the relevant experimental population is key to the design of a successful multi-level experiment.

If it is not possible to identify these social structures, or if it is not possible to conduct an experiment which incorporates variation in these structures, then the researcher needs to think carefully about the type of inference which is drawn from the analysis of this data. Inferences drawn from experimental data which have not explicitly incorporated spillovers but where contagion of the treatment is likely may have greater problems with external validity – the same treatment, administered in a different social context, is unlikely to generate the same effect. However, if inferences are presented at the group level where the potential for spillovers is acknowledged but not incorporated into the experimental design, then in other contexts where the social structures are similar the treatment effects should be similar. Thus the caveat for the researcher should be to acknowledge the potential concern with external validity and to present the group-level treatment effects at the level above which the social interactions have occurred.

Figure 1 Goes Here

In our example, we randomly assign groups in the population to treatment and control. Within the groups assigned to treatment, we then again repeat the random assign-

ment, randomly assigning households to treatment and control. Finally, within the households assigned to treatment, we again repeat the random assignment, randomly assigning individuals to treatment and control. Treatment is administered at the individual level, to all members of the third stage of the randomization who are assigned to treatment. We conduct randomization at each level where we anticipate social interactions will take place. McConnell, Sinclair and Green (2009) follow this recommendation for experimental analysis in Los Angeles and Chicago. This process allows for the identification of both the spillover effects and the true treatment effect while removing the potential bias associated with the spillover. The SUTVA violations have become a quantity of interest, allowing for inferences on interpersonal interactions.

## **6 Empirical Tests of Spillovers and Estimation of the Intent-to-Treat Effect**

Here we provide two sets of recommendations for estimation of the intent-to-treat effect (ITT). The first set of recommendations are strategies for situations where there are inconsistencies between the random assignment and treatment. These are statistical corrections in the absence of design approaches. The second set of recommendations are strategies for situations where the experimental design has explicitly adjusted for spillover. Both sets of strategies are simple to adopt and require minimal assumptions.

Under the first condition where there exist inconsistencies between the random assignment and the administration of the treatment, we recommend two strategies for analysis. Suppose that an experiment has been conducted where the randomization occurred at the individual level but the treatment was administered at the group level. In this case, we first recommend clustering the standard errors at the group level when estimating the ITT. This clustering explicitly acknowledges the correlation that is likely to exist within the group as a result of the administration of the treatment and adjusts for the lack of independence of all observations within the group (Green and Vavreck 2008). Note, however, that this adjustment is not sufficient to account for the potential biases that have

occurred as a result of social interactions. Correlation in the standard errors does not account for the possibility that individuals who are assigned to treatment, for example, may have been indirectly treated multiple times from other individuals assigned to treatment nor the possibility that individuals assigned to control may have been indirectly treated from individuals assigned to treatment. Our second recommendation, if there are a sufficient number of group-level observations, is to conduct analysis either via a hierarchical linear model or via meta-analysis. Meta-analysis, for example, does not require homoskedasticity across groups, which is a necessary assumption with the inclusion of group-level fixed-effects, which assumes that there is a group-specific effect (Gerber et al 2008, Sinclair, McConnell and Michelson 2009). While this is more often a concern when there are failure-to-treat instances, there is still likely to be group-level variation which is not properly accounted for via fixed-effects.

Under the second condition where the experiment has incorporated the multi-level context into the experimental design, we again recommend two strategies for analysis. The first allows us to identify the presence of any social spillovers within our explicitly specified social contexts. We recommend using Hansen's J-statistic to determine if there is over-identification in the outcome behavior of the individuals in our sample when using each of the levels of random assignment as instrumental variables. This allows us to evaluate whether or not there is enough evidence to reject the null that, for example, both the individual and group level assignment variables are valid instruments. If we can reject this null hypothesis, then this suggests that in fact we do not have spillovers – a clear test which then implies that we do not need to incorporate the multi-level structure in any additional analyses. If we cannot reject this null, then we recommend a second strategy for analysis. In the estimation of the ITT, it is then necessary to incorporate parameters to separately estimate the effects of each level of assignment. A visual demonstration for each of these estimates is included in Table 1.

Table 1 Goes Here

That is, consistent with our example, we then need to estimate the effect of having

been assigned to group-level treatment but not household-level nor individual-level treatment compared to the group-level control. We also estimate the effect of having been assigned to household-level treatment but not individual-level treatment compared to household-level control and finally the effect of having been assigned to individual-level treatment compared to individual-level control. In Table 1, the difference between the two columns produces the quantity of interest. If either of the first two rows produce statistically significant effects, then the final row is likely to be biased. Thus, an estimate of the true treatment effect would subtract each of these spillover effects from the final row to estimate the true ITT.

## **6.1 Limitations and Best Practices**

There are two limitations which emerge from conducting multi-level experiments. The first is simply the challenge in identifying the levels at which social interactions are likely to take place. Given that many experiments may take advantage of geography, however, within which to conduct the experiment, this may not be a large hurdle for many types of experiments. The second limitation which emerges is that the construction of a multi-level experiment may require a larger experimental population in order to have sufficient statistical power to identify the spillover effects. Thus the additional limitation is the loss of efficiency which emerges with the construction of multiple control groups. In many populations this is not a concern, as there are sufficiently many individuals that the addition of additional control group members does not limit the feasibility of the experiment. Yet there may be contexts within which either it is difficult to locate additional control group participants or where the requirement of additional control groups which are geographically dispersed increases the cost of conducting the experiment.

The best practice when the experiment is limited by the potential size of the control group and thus cannot be reasonably conducted on a multi-level scale is to present the group-level effects and to acknowledge the potential presence of spillovers. Spillovers can take on multiple forms; some individuals may be able to be treated multiple times,

and it is possible that both individuals assigned to treatment and control may be subject to spillovers. This makes it difficult to calculate the appropriate counterfactuals for how large of a potential spillover effect could have been present with observed experimental data. Given the lack of knowledge about the direction of the bias the spillovers could take, it is impossible to ex-ante predict the effects of potential SUTVA violations. Under certain kinds of spillovers, the estimates could in fact converge to the actual true value of the treatment effect, for example.

One final limitation of the multilevel context, albeit more applicable where there are failure-to-treat cases, is that the bias and loss of efficiency from using instrumental variables when the contacts are made uniformly across all groups is different than when the contacts are concentrated in some groups, as is the potential case in the multi-level context.

## **7 Conclusion: Recommendations for Experimental Design and Future Research**

We recommend that in multi-level contexts, that randomization take place not only at the individual level but also at all appropriate social structure levels. If treatment is then administered at the individual level, it is clear how to draw inferences about spillovers, allowing us great insight into the way in which politics can be socially transmitted and the role of interpersonal influence. In this sense, SUTVA violations have become a quantity of interest.

Most importantly, however, is the role of transparency in multi-level contexts. To the extent that it is possible, researchers must document whether their experimental treatments are administered at a group level and acknowledge that these strategies require shifts in their estimation procedures. Furthermore, if researchers believe that their experiment is operating within a multi-level context, it is key that this be documented so that future researchers can incorporate these facts into future experiments.

While most of this essay has been most applicable for field experiments, it is possible

that the multi-level context is applicable for alternative situations as well. For example, in laboratory settings, there may be existing social structures which complicate the analysis if the participants do not interact in a way which preserves their anonymity. Undergraduate university participants may belong to the same fraternity or sorority. Laboratory experiments also need to be careful about inferences drawn after randomization, as while ex-ante the groups are identical in terms of their covariate distributions, with small numbers it is increasingly likely to have differences in distributions in terms of either the observed or unobserved characteristics.

There are many situations where experiments are conducted within multi-level settings. Voter mobilization experiments, where treatment and randomization occur at the level of the individual, are clearly prey to potential SUTVA violations. Within existing social and political organizations, there may be hierarchical or geographically-distributed groups that are also subject to potential SUTVA violations, such as statewide organizations with local chapters and individual members. Clearly these concerns are relevant for experiments conducted on college campuses, where individual participants may be residents in the same dorm, for example.

Experimenters need to be sure to consider failure-to-treat situations and the ways in which they may further complicate these analyses. This essay has not dealt explicitly with failure-to-treat, but these instances require additional assumptions under which to draw inferences. In particular, many kinds of analyses use the random assignment variable as an instrumental variable in order to estimate the treatment-on-treated effect (TOT). This approach is not appropriate in cases where there is social transmission of treatment within the experimental population as the assignment variable fails to capture the indirect treatment.

Multi-level experiments have the potential to yield great insights into the ways in which humans interact; with careful experimental design, the SUTVA violations have the potential to open up new avenues of research previously reliant upon heroic assumptions. Each additional randomization does not add to the technical difficulty of implementing

the experiment, as it is still possible for the experimental design to include the same number of participants assigned to be administered the treatment. It is the addition of the new control groups that allows for the identification of the spillover effects. Researchers should be aware of the statistical power necessary to test each of these spillover categories, however, and should follow recommendations in McConnell, Sinclair and Green (2009) to adjust the size of their experimental population accordingly.

This methodological improvement has the potential to encourage different kinds of inferences in randomized field experiments. This is also a technique that allows the discovery of supportive evidence for individuals who study network analysis via survey data to understand social structure. This method can also be extended to include additional randomizations to study spillover in many directions – for example, we could also include a category where we compared individuals assigned to treatment who were paired with control to individuals assigned to treatment who were paired with treatment to individuals assigned to control – this would allow us to see if in fact the pairing of treated-with-treated would increase the effect of the treatment as well. SUTVA violations have the potential to be extremely interesting quantities of interest. As we develop new and interesting ways to measure spillovers, these quantities will allow us to inform which types of theories are most applicable in the social transmission of politics. We do not yet know whether or not the instigation of political behavior is due to generated conversations, heightened interest, or persuasion. The measurement of spillover will offer one set of illustrations for where our theories should focus.

## 8 Works Cited

Adato, M., Coady, D., and Ruel, M. 2000. "An operations evaluation of PROGRESA from the perspective of beneficiaries, promotoras, school directors and health staff." Final report, *International Food Policy Research Institute*, 2033 K Street, NW Washington, DC 20006.

Arceneaux, Kevin. 2005. "Using Cluster Randomized Field Experiments to Study Voting Behavior", *The Annals of the American Academy of Political and Social Science* Vol. 601, No. 1, 169-179.

Besley, Timothy and Anne Case. "Modeling Technology Adoption in Developing Countries", *American Economic Review*, 83 (2), 1993, pp. 396-402.

Berelson, Lazarsfeld and McPhee. 1954. "Voting". Chicago, University of Chicago Press, pp. 168-175.

Brock, W.A. and S.N. Durlauf. 1999. "A formal model of theory choice in science". *Economic Theory* 14: 113-130.

Cacioppo, John T., James H. Fowler and Nicholas A. Christakis. 2009. "Alone in the Crowd: The Structure and Spread of Loneliness in a Large Social Network". *Journal of Personality and Social Psychology*, forthcoming.

Coleman, J. E. Katz and H. Menzel. (1966). *Medical Innovation*. Bobbs-Merrill, Indianapolis, IN.

Conley, TG and CR Udry, 2009. "Learning About a New Technology: Pineapple in Ghana". Forthcoming *American Economic Review*.

DeMarzo, P. D. Vayanos and J. Zwiebel. 2003. "Persuasion Bias, Social Influence and Unidimensional Opinions". *Quarterly Journal of Economics*, Vol. 118, No. 3, pp. 909-968.

Duflo, Esther, Michael Kremer and Jonathan Robinson. "Understanding Technology Adoption: Fertilizer in Western Kenya". Working Paper, 2006.

Fowler, James H. and Nicholas A. Christakis. 2008. "Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study". *British Medical Journal* 337: a2338.

Gerber, Alan S., Donald P. Green and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment". *American Political Science Review* 94: 653-663.

Green, Donald P., Alan S. Gerber, and David W. Nickerson. 2003. "Getting Out the Vote in Local Elections: Results from Six Door-to-Door Canvassing Experiments," *Journal of Politics*, Vol. 65:1083-96.

Green, Donald P. and Lynn Vavreck. 2008. "Analysis of Cluster-Randomized Experiments: A Comparison of Alternative Estimation Approaches," *Political Analysis*, 16(2): 138-152.

Gerber, Alan S. and Donald P. Green. 2008. "Get Out the Vote!" Brookings Institution Press, second edition, Washington DC.

Hansen, Ben B. and Jake Bowers. 2008. "Attributing Effects to a Cluster Randomized Get-Out-The-Vote Campaign". *Journal of the American Statistical Association*. Forthcoming.

Huckfeldt, Robert, Paul E. Johnson and John Sprague. 2004. *Political Disagreement* Cambridge University Press, Cambridge.

Imai, Kosuke, Gary King, and Clayton Nall. 2009. "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation." *Statistical Science*, Forthcoming.

King, Gary, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T. Moore, Jason Lakin, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila and Héctor Hernández Llamas. 2007. "A 'Politically Robust' Experimental Design

for Public Policy Evaluation, with Application to the Mexican Universal Health Insurance Program". *Journal of Policy Analysis and Management* 26(3):479 - 509.

Krueger, A. 1999. "Experimental estimates of education production functions". *Quarterly Journal of Economics*, 114, 497 – 532.

Manski, Charles. "Identification of Exogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, 60, 1993, pp. 531-542.

McConnell, Margaret A., Betsy Sinclair and Donald P. Green. "Detecting Spillover in Social Networks: Design and Analysis of multi-level Experiments". Working paper.

McClurg, Scott. 2006. "Political Disagreement in Context: The Conditional Effect of Neighborhood Context, Discussion, and Disagreement on Electoral Participation." *Political Behavior* 28(4), pp. 349- 66.

Miguel, E. and M. Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities". *Econometrica*, Vol. 72, No. 1, pp. 159-217.

Munshi, Kaivan. "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution", *Journal of Development Economics*, 73(1), 2004, pp. 185-213.

Nair, Harikesh, Puneet Manchanda, and Tulikaa Bhatia. (2008) "Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders". SSRN 937021, Working paper.

Nickerson, David. "Is Voting Contagious? Evidence from Two Field Experiments", *American Political Science Review*, Forthcoming 2008.

Rubin, Donald B. 1980. "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment". *Journal of the American Statistical Association*, Vol. 57, No. 371 (Sep.), pp. 591-593.

Rubin, Donald B. 1990. "Formal Modes of Statistical Inference for Causal Effects". *Journal of Statistical Planning and Inference* 25: 279-292.

Panagopoulos, Costas and Donald P. Green. 2006, "The Impact of Radio Advertisements on Voter Turnout and Electoral Competition." Paper prepared for presentation at the Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 20-23.

Sinclair, Betsy and Brian Rogers. 2009. "Political Networks: The Relationship Between Candidate Platform Positions and Constituency Communication Structures". Working paper.

Sinclair, Betsy, Margaret McConnell and Melissa R. Michelson. 2009. "Strangers vs Neighbors: The Efficacy of Grassroots Voter Mobilization". Working paper.

## 9 Tables and Figures

Figure 1: Multi-level Experiment Design

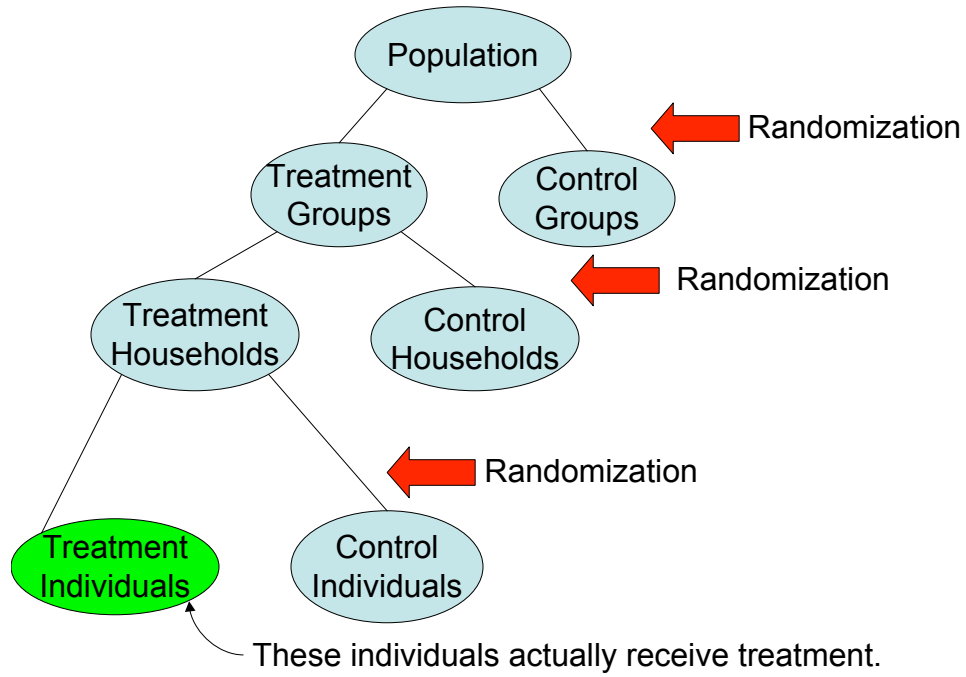


Table 1: ITT Effects

Assignment (Group, Household, Individual)	Assignment (Group, Household, Individual)	Quantity of Interest
Control, Control, Control	Treatment, Control, Control	Group-level Spillover Household-level Spillover Treatment Effect (Potentially Biased)
Treatment, Control, Control	Treatment, Treatment, Control	
Treatment, Treatment, Control	Treatment, Treatment, Treatment	