Journal of Experimental Political Science

http://journals.cambridge.org/XPS

Additional services for ***Journal of Experimental Political Science:***

Email alerts: Click here
Subscriptions: Click here
Commercial reprints: Click here
Terms of use : Click here

Information Spillovers: Another Look at Experimental Estimates of Legislator Responsiveness

Alexander Coppock

Journal of Experimental Political Science / Volume 1 / Issue 02 / December 2014, pp 159 - 169
DOI: 10.1017/xps.2014.9, Published online: 12 January 2015

Link to this article: http://journals.cambridge.org/abstract_S2052263014000098

How to cite this article:
Alexander Coppock (2014). Information Spillovers: Another Look at Experimental Estimates of Legislator Responsiveness. Journal of Experimental Political Science, 1, pp 159-169 doi:10.1017/xps.2014.9

Request Permissions : Click here

CAMBRIDGE JOURNALS

Downloaded from http://journals.cambridge.org/XPS, IP address: 128.59.222.12 on 12 Jan 2015

Journal of Experimental Political Science **1** (2014) 159–169
doi:10.1017/xps.2014.9

Information Spillovers: Another Look at Experimental Estimates of Legislator Responsiveness

Alexander Coppock*

Abstract

A field experiment carried out by Butler and Nickerson (Butler, D. M., and Nickerson, D. W. (2011). Can learning constituency opinion affect how legislators vote? Results from a field experiment. *Quarterly Journal of Political Science* 6, 55–83) shows that New Mexico legislators changed their voting decisions upon receiving reports of their constituents' preferences. The analysis of the experiment did not account for the possibility that legislators may share information, potentially resulting in spillover effects. Working within the analytic framework proposed by Bowers et al. (2013), I find evidence of spillovers, and present estimates of direct and indirect treatment effects. The total causal effect of the experimental intervention appears to be twice as large as reported originally.

Keywords: Field experiment, spillovers.

INTRODUCTION

Butler and Nickerson (2011) report the results of an innovative field experiment testing the responsiveness of legislators to public opinion in New Mexico. Most previous studies of responsiveness note a positive correlation between public opinion and legislators' choices, which may be due to electoral concerns, the similarity of preferences, or public responsiveness to elite opinion, among many other possible explanations. Butler and Nickerson isolate a single causal channel—the effect of *learning* public opinion on legislators' voting decisions—by randomly providing some legislators with survey measures of their constituents' preferences. The headline finding from their study is that representatives change their voting behavior upon acquiring novel public opinion information.

The estimates of responsiveness recovered by Butler and Nickerson (2011) rely on an assumption of non-interference (Cox 1958; Rubin 1980): Legislators respond

The author is grateful to Donald P. Green, Robert Erikson, Gregory Wawro, Peter Aronow, Lindsay Dolan, Albert Fang, and two anonymous reviewers for helpful comments and suggestions, and to Daniel Butler and David Nickerson for providing replication materials.
*Columbia University, New York, NY, USA; e-mail: ac3242@columbia.edu

© The Experimental Research Section of the American Political Science Association 2015

only to their own treatment status and not to the treatment status of others. This assumption requires that legislators not share treatment information with one another, which is at odds with the observation by Kingdon (1973, p. 6) that legislatures are information-sharing networks. As such, a straightforward comparison of treatment and control groups may not recover the estimand of interest.

The standard estimand is the average effect of public opinion information on an individual legislator's voting behavior. Consider two possible treatment assignments—one in which zero legislators are treated, and another in which exactly one is treated. The difference in behavior for the treated legislator across the two treatment assignments is her individual treatment effect. The average of all such individual treatment effects is the standard average treatment effect (ATE). Now consider two other possible treatment assignments—one in which all legislators are treated, and another in which exactly one is untreated. Again, the difference in behavior for the untreated legislator is an individual treatment effect. Under the non-interference assumption, this individual treatment effect is the same as the one described in the first scenario. If, however, information travels from the treated legislators to the single untreated legislator, then this individual treatment effect and the resulting ATE may be different.

Finally, consider the random assignment that did occur: 35 of 70 legislators were sent information. The difference-in-means is an estimate of an ATE. Under the non-interference assumption, it is an estimate of the ATE that would obtain regardless of the total number of legislators treated. If this assumption does not hold, however, it is unclear *which* ATE is being estimated.[1] Strictly speaking, the original analysis estimated the average effect of being treated versus untreated while half the legislature is treated. Nevertheless, much can still be learned from the random assignment of information to legislators. The social and political relationships linking treated legislators to their colleagues dictate the extent to which information sharing occurs. The random assignment therefore governs the level of information to which each legislator is exposed, providing a basis for inference for both direct and indirect effects.

EXPERIMENTAL CONTEXT

The experiment was carried out during a special session of the New Mexico state legislature in 2008. The state had projected a budget surplus, and the Senate Bill 24 (SB24) proposed to return the surplus to taxpayers in the form of a rebate. The bill was understandably popular among New Mexicans, but the source of the surplus, oil revenues, had shrunk considerably due to dropping oil and gas prices.

[1]See Hudgens and Halloran (2008) and Sobel (2006) for fuller discussions of causal estimands under interference.

This circumstance put a wedge between constituents favoring a tax rebate and those more concerned with the fiscal impacts of the decline in oil revenues. Butler and Nickerson (2011) shared survey estimates of support for the rebate with 35 of the 70 state representatives using matched-pair randomization. Treated legislators were mailed letters containing district-specific information: Some legislators found out that their districts had high levels of support for the rebate, while others found out that the districts had low levels of support.

Statewide, SB24 was a relatively popular measure, and in the absence of additional information, legislators may have assumed that the bill was popular in their own home districts as well. Butler and Nickerson suspected that legislators would expect high levels of support for the rebate and would be surprised by low levels. Therefore, there is a class of legislators for whom they expected to see no effect of the information treatment on vote choice: those whose districts *do* in fact support the popular bill. Legislators whose districts have low levels of support are the ones whose voting behavior might change as a result of treatment. The average effect of treatment among all legislators was negative but statistically insignificant. The average effect of treatment among legislators from districts with low support for spending was negative, statistically significant, and constitutes the major finding, as it provides support for the idea that those legislators who gained "counterintuitive" information from the public opinion data were the ones whose vote choice was affected.

In order to reanalyze the experimental results in a framework that allows for spillovers, we must have a sense of how the legislature is organized. Squire (2007) notes the low level of professionalization in the New Mexico state legislature: If any sharing of treatment information did occur, it probably took place via the personal relationships between individual legislators. Below, I will assume the closeness of those relationships determines the level of information shared.

METHOD

In this section, I describe the measures used to estimate the strength of the personal connections between legislators and the hypothesis-testing framework for assessing treatment effects.

Modeling Information Spillovers

The random assignment of public opinion information to legislators introduced 35 treatments into the information sharing network of the New Mexico legislature. In order to estimate their impact, we need a function that maps each random assignment onto a set of potential outcomes. The most flexible mapping would associate each of the $2^{35} = 3.4 \times 10^{10}$ possible random assignments with a unique

potential outcome for each legislator. [2] The most parsimonious function (that still serves a scientific purpose) would map each of the 30 trillion random assignments onto just two potential outcomes for each unit—a treated and an untreated potential outcome, ignoring the assignment of the other units entirely.[3] The approach taken here is a middle ground between these extremes. The potential outcome revealed by legislator i is a function of his or her own treatment assignment z_i as well as a weighted average of all the other legislators' treatment assignments. The weights given to the other legislators' treatment assignments are determined by an information network model, the choice of which must be theoretically driven.

This model of potential outcomes has intuitive appeal: Legislators are indirectly exposed to more or less information, depending on *which* other legislators are assigned to direct treatment. As more of a legislator's close colleagues are treated, the legislator will receive "more" spillovers. The problem is now computationally tractable because the high-dimensional potential outcome space has been reduced to just two dimensions — direct and indirect exposure to treatment.[4]

The investigation of indirect effects requires a model of the pathways along which spillovers can occur. The model developed here[5] is based on the ideological similarity of legislators. Estimates of legislator ideology are drawn from an analysis of roll call votes using the nominate package in R (Poole et al., 2011). The voting data are based on 17 key votes (as chosen by Vote Smart (2008)) from the regular legislative session directly preceding the 2008 special session. Focusing only on the first dimension, each legislator is assigned a W-NOMINATE ideology score between –1 and 1. In order to then calculate the similarity between any two legislators i and j, the following formula was employed:

$$\text{Similarity}_{i,j} = \frac{2 - |Ideo_i - Ideo_j|}{2} \tag{1}$$

For ideology values between -1 and 1, Equation (1) varies between 0 and 1. The $n \times n$ matrix of similarity scores is denoted $\mathbf{\Gamma}$. Legislators are "exposed" to treatment via every treated colleague in the chamber, but exposure is higher if the treated colleagues are more similar. The formula for the raw level of exposure is given in Equation (2), where z_j is the treatment assignment of legislator j,

$$\text{Raw Exposure}_i = \sum_{j=1}^{n} \text{Similarity}_{i,j} \times z_j, j \neq i \tag{2}$$

Legislator i's *Raw Exposure* is the weighted average of the other legislators' treatment assignments, where the weights are the similarity scores between legislator i

[2]Recall that Butler and Nickerson (2011) employed matched-pair random assignment in order to reduce the variability of their estimates.

[3]See Manski (2013) for a discussion of mapping random assignments to potential outcomes.

[4]For a discussion of reducing spillovers to a scalar quantity, see Hong and Raudenbush (2006, p. 902).

[5]Please consult the online appendix for an analysis based on the geographic adjacency of legislative districts as well as alternative parameterizations of ideological closeness.

and colleagues. However, two complications are introduced by this specification of spillovers. First, legislators have different probabilities of experiencing a given level of exposure. Legislators whose set of similarity scores are higher than average are more likely to experience spillovers. Left uncorrected, this would mean that the exposure variable would be correlated with ideology and associated unobservable characteristics. Second, *Raw Exposure* is mechanically correlated with direct treatment: when legislator i is treated, only 34 of the remaining 69 legislators can be treated, resulting in a lower level of raw exposure.

Accounting for expected exposure addresses both complications. Expected exposure is calculated by simulating what exposure would have been under a large set possible randomizations,[6] indexed by k for randomizations in which legislator i is in treatment and by l when i is in control:

$$\text{Expected Exposure}_{i,z_i=1} = \frac{\sum_{k=1}^{K}\sum_{j=1}^{n} \text{Similarity}_{i,j} \times \text{z}_{j,k}}{K}, \quad j \neq i, z_{i,k} = 0 \tag{3}$$

$$\text{Expected Exposure}_{i,z_i=0} = \frac{\sum_{l=1}^{L}\sum_{j=1}^{n} \text{Similarity}_{i,j} \times \text{z}_{j,l}}{L}, \quad j \neq i, z_{i,l} = 1 \tag{4}$$

The variable of interest, then, is not the raw level of exposure, but the difference between the raw level and the expected level, which we will call *Net Exposure*. This variable is no longer related to network position or direct treatment assignment,

$$\text{Net Exposure}_i = \begin{cases} \text{Raw Exposure}_i - \text{Expected Exposure}_{i,Z_i=1}, & \text{if } Z_i = 1 \\ \text{Raw Exposure}_i - \text{Expected Exposure}_{i,Z_i=0}, & \text{if } Z_i = 0 \end{cases} \tag{5}$$

Finally, in order to ease the interpretation of the *Net Exposure* variable, it is standardized by the z-score transformation. The coefficients on the indirect treatment variable can be interpreted as the change in probability of voting yea due to a 1 standard deviation increase in net exposure.

To summarize the model of spillovers introduced here: Beginning with a similarity matrix $\mathbf{\Gamma}$ and a random assignment $\mathbf{z}$, we calculate a set of raw exposures according to Equation (2). We then subtract off the expected level of exposure for each unit, by treatment condition, according to Equations (3) and (4). Finally, we standardize the resulting variable. Collectively, these operations are denoted as $g(\mathbf{\Gamma} z)$.

Hypothesis Testing

Building on the work of Rosenbaum (2002, 2007), Hodges and Lehmann (1963), and Fisher (1935), Bowers et al. (2013) offer a framework for evaluating hypotheses given a causal model. Other analytic frameworks for spillovers such as the one proposed by Aronow and Samii (2013) based on inverse probability weighting could be used, although the estimation of assignment probabilities would be a

[6] All simulations followed the block-randomized procedure used to generate the actual randomization that Butler and Nickerson (2011) carried out.

computational challenge due to the assumed spillover structure. The approach advocated by Bowers et al. (2013) has the intuitive appeal of associating a p-value with sets of hypothesized parameter values. Confidence regions are constructed by observing the parameter sets associated with p-values exceeding some α-level.

A causal model is a function $\mathcal{H}(y_{i,\mathbf{z}}, \mathbf{w}, \theta) = y_{i,\mathbf{w}}$ that translates outcomes observed under treatment assignment vector $\mathbf{z}$ to outcomes that would be observed under treatment assignment vector $\mathbf{w}$. The parameter vector θ is the collection of causal effects of treatment. By comparing observed data with the data that would be generated by the model, we can calculate p-values associated with hypothesized values of θ. The most common hypothesis, the sharp null hypothesis, is that the values of θ are equal to zero for all legislators. Fisherian significance tests ask, supposing the sharp null hypothesis of no effect were true, how frequently would we observe a test statistic (such as the difference-in-means) as large or larger in magnitude as observed in the experimental data among all possible random assignments. The same logic can be extended to any sharp hypothesis in which a constant effect is proposed to hold for all subjects.

In order to associate p-values with any hypothesized values for θ, we must employ the notion of a "uniformity trial" (Rosenbaum 2007), which is the vector of outcomes that would have occurred if no treatment had been administered to any unit, written as $\mathbf{y_0}$. The uniformity trial for the spillover model investigated here is given in Equation (6). Note that this equation is an algebraic manipulation of the causal model of the form $\mathcal{H}(y_{i,\mathbf{z}}, \mathbf{w}, \theta) = y_{i,\mathbf{w}}$, where β_1 and β_2 are the parameters in θ,

$$y_{i,\mathbf{0}} = y_{i,\mathbf{z}} - \beta_1 z_i - \beta_2 g(\mathbf{\Gamma} z) \tag{6}$$

The uniformity trial is calculated by subtracting off the hypothetical direct and indirect treatment effects, β_1 and β_2, from the observed outcomes under treatment assignment $\mathbf{z}$. The function $g(\cdot)$, defined in section "Modeling Information Spillovers," translates assignment $\mathbf{z}$ and the information network $\mathbf{\Gamma}$ into a scalar-valued quantity that captures the amount of spillover received by each unit.

For each simulated random assignment $\mathbf{w}$, treatment effects are added back to the uniformity trial according to the causal model implicit in Equation (6). Following advice given in Aronow (2013), I use the sum of squared residuals (SSR) from an Ordinary Least Squares (OLS) regression of hypothesized outcomes on direct and indirect treatment vectors as a test statistic. The simulated statistics will be large relative to the observed statistic when the hypothesized values for β_1 and β_2 are improbable. The p-value associated with each hypothesis is the fraction of simulations in which the observed SSR statistic exceeds the simulated SSR statistic. Bowers et al. (2013) suggest choosing the pair with the highest p-value as an approximation to the Hodges–Lehmann point estimate (Hodges and Lehmann 1963), which is the approach taken here.

RESULTS

This section will present results under two interference assumptions. The first analysis makes the standard non-interference assumption, i.e., $\mathbf{\Gamma}$ is a matrix of zeros and units' potential outcomes respond only to their own treatment assignments. The second set of results will assume that information can travel over the ideological similarity network. A summary table of all results presented below is available in the online appendix.

Replication of Original Analysis

Panel A in Figure 1 replicates the original analysis (i.e., supposing the indirect effects are exactly zero) in the hypothesis-testing framework discussed above. Proposed values for the direct effect of treatment are displayed on the x-axis and simulated p-values on the y-axis. The points are colored according to their p-values in order to keep the visual display similar to the 2-parameter plots to come. The p-value is maximized at –0.16, identical to the difference-in-means estimate. As the hypothetical values for ATEs diverge from –0.16, the p-values decrease. A hypothesis of interest is the hypothesis that the true parameter is equal to zero—the corresponding p is slightly higher than 0.05, indicating that this result is not statistically significant at conventional levels. The 95% confidence region extends from -0.34 to 0.02.

Panel B in Figure 1 shows the heterogeneous effects of treatment. Treatment effect hypotheses in high-support districts are plotted on the x-axis, and hypotheses for low-support districts are plotted on the y-axis. The p-value associated with each pair of hypotheses is indicated by the color scale: darker shaded squares indicate higher p-values. For this and all future plots, the 95% confidence region is the area with any shading at all. The p-value is maximized at $(0.05, -0.37)$, the treatment effects in high-support and low-support districts, respectively. As in the original analysis, the confidence region is bounded away from zero for the low-support districts only.

Ideological Similarity Spillover Model

The results presented in Figure 2 relax the assumption of no indirect effects. Direct effect (β_1) hypotheses are presented on the x-axis and indirect effect (β_2) hypotheses are presented on the y-axis. The highest p-value point is $(-0.300, -0.225)$, constituting the point estimates of direct and indirect effects. Legislators who received treatment letters were 30 percentage points less likely to vote for the bill. Legislators who were exposed to a 1 standard deviation increase in second-hand exposure were 22.5 percentage points less likely to vote yea. The 95% confidence region does not cross zero for either the direct or indirect effect, indicating that these results are statistically significant. These results provide evidence that legislators whose close colleagues received treatment were less likely to vote for the bill.

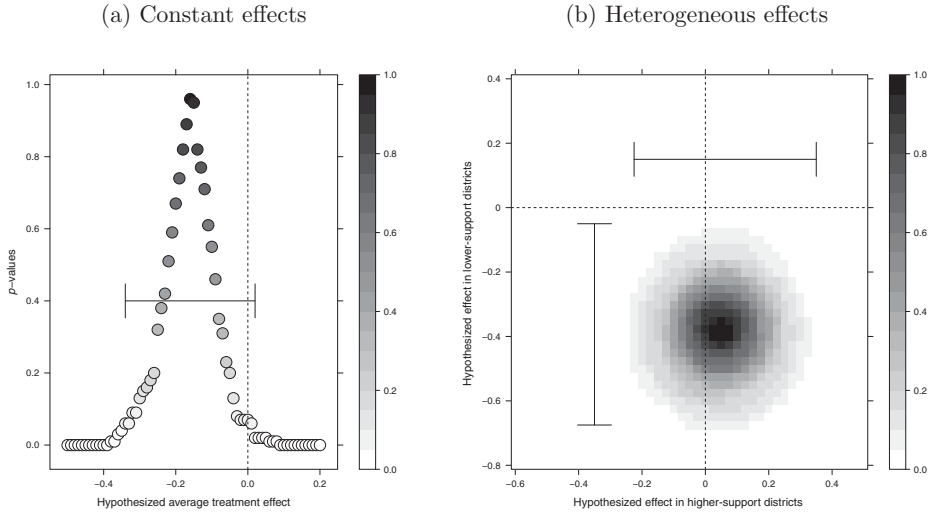

Figure 1
p-Value Maps for Models Assuming No Spillover

Figure 3 displays the results of a model analogous to the heterogeneous effects model presented above, but allowing for spillovers over the ideological similarity network.[7] Panel A presents a p-value map in the low-support districts. The highest p-value point is $(-0.550, -0.125)$. The direct effect estimate in low-support districts is, as above, highly statistically significant: no hypotheses in which the proposed value for the direct effect is above zero receive positive probability. The indirect effect appears to be statistically significant as well—the 95% confidence region extends up to, but not across, the zero line. In Panel B, the highest p-value point is $(-0.175, -0.300)$. As in the no spillovers model, the direct effect of treatment cannot be distinguished from zero in the high-support districts. The indirect effect, however, is strongly negative and the 95% confidence region is bounded away from zero. Allowing for the possibility of spillover effects does not alter the finding that direct treatment effects appear to be stronger among legislators from low-support districts than from high-support districts.

CONCLUSION

This paper has explored ways in which interference between units could have occurred in the Butler and Nickerson (2011) experiment. The information network is not observed directly, so modeling assumptions have to be made. Strongly suggestive

[7]The causal model in Equation (6) is investigated for high- and low-support districts separately. See the online appendix for a discussion of generalization of the model that allows for the joint estimation of all four parameters.

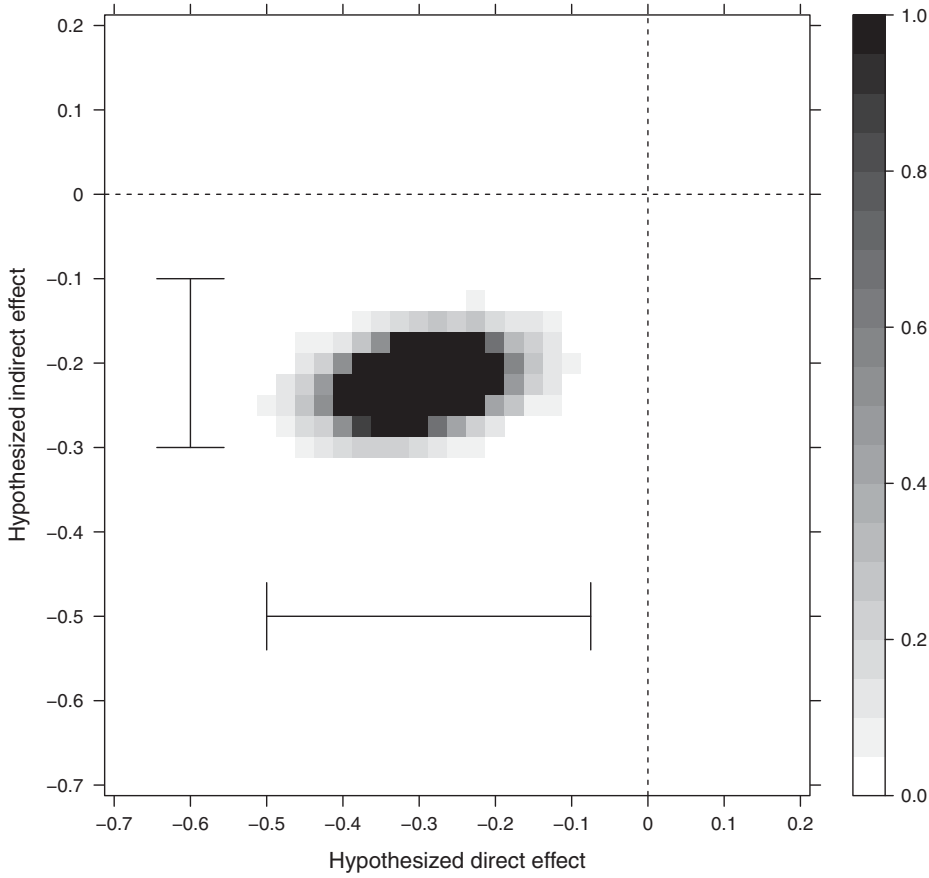

Figure 2
***p*-Value Map of Ideological Similarity Spillover Model**

evidence of spillover was uncovered using a measure of ideological similarity of legislators. Exposure to the policy preferences of constituents—either one's own or those of colleagues—decreased legislators' probability of voting for SB24. Using the difference-in-means estimate, the treatment letters appear to have changed five and a half votes ($35 \times -0.16 = -5.6$). The similarity network model estimates that in the absence of treatment, 61 legislators would have voted for the bill, whereas 51 actually did. The estimated total causal effect is therefore 10 votes.

Experimental studies of information treatments are especially prone to non-interference violations. Within political science, many areas of study are concerned with the impact of information, including investigations of corruption, persuasion, and identity. Interference can pose a major challenge for estimation; however, the specification of theoretically driven spillover models may provide at least a partial solution. The bottom line for the substantive results of the Butler and Nickerson

(a) Among low-support districts. (b) Among high-support districts.

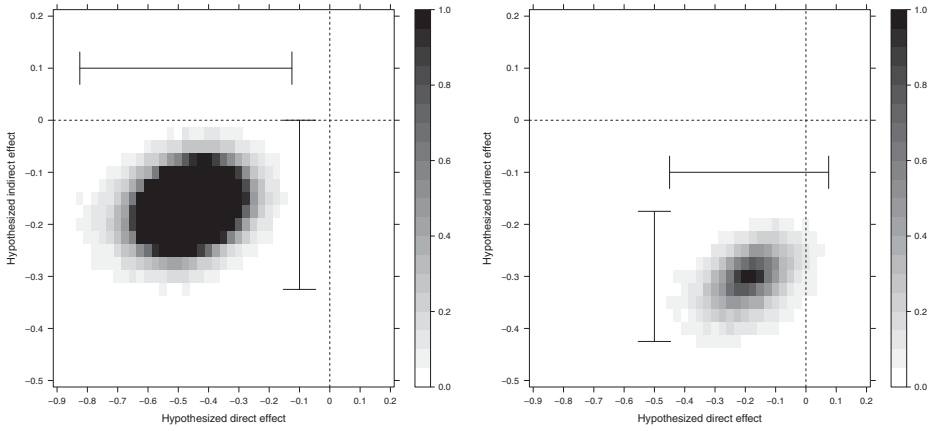

Figure 3
***p*-Value Maps by Level of Support**

(2011) experiment has not changed—if anything, the magnitude of the total causal effect has nearly doubled. Legislators respond to public opinion information, whether that information comes in the form of letters sent to their offices or in the form of a colleague sharing news.

SUPPLEMENTARY MATERIAL

To view the supplementary material for this paper, please visit http://dx.doi.org/10.1017/XPS.2014.9.

REFERENCES

Aronow, P. M. 2013. "Comment: Reasoning About Interference Between Units". Unpublished Mm, 1–7.

Aronow, P. M., and Samii, C. 2013. *Estimating Average Causal Effects Under Interference Between Units*. Unpublished manuscript, Yale University, New Haven, CT.

Berger, R. L., and Boos, D. D. 1994. *P* Values Maximized over a Confidence Set for the Nuisance Parameter. *Journal of the American Statistical Association* 89 (427): 1012–16.

Bowers, J., Fredrickson, M. M., and Panagopoulos, C. 2013. Reasoning About Interference Between Units: A General Framework. *Political Analysis* 21 (1): 97–124.

Butler, D. M., and Nickerson, D. W. 2011. Can Learning Constituency Opinion Affect how Legislators Vote? Results from a Field Experiment. *Quarterly Journal of Political Science* 6: 55–83.

Cox, D. R. 1958. *Planning of Experiments.* New York: Wiley.

Fisher, R. A. 1935. *The Design of Experiments.* Edinburgh: Oliver and Boyd.

Hodges, J. L., Jr., and Lehmann, E. L. 1963. Estimates of Location Based on Rank Tests. *The Annals of Mathematical Statistics* 34 (2): 598–611.

Hong, G., and Raudenbush, S. W. 2006. Evaluating Kindergarten Retention Policy. *Journal of the American Statistical Association* 101 (475): 901–10.

Hudgens, M. G., and Halloran, M. E. 2008. Toward Causal Inference with Interference. *Journal of the American Statistical Association* 103 (482): 832–42.

Jacobs, L. R., Lawrence, E. D., Shapiro, R. Y., and Smith, S. S. 1998. Congressional Leadership of Public Opinion. *Political Science Quarterly* 113 (1): 21–41.

Kingdon, J. W. 1989. *Congressman's Voting Decisions*. Ann Arbor, MI: University of Michigan Press.

Manski, C. F. 2013. Identification of Treatment Response with Social Interactions. *The Econometrics Journal* 16 (1): S1–23.

Nolen, T. L., and Hudgens, M. G. 2011. Randomization-Based Inference Within Principal Strata. *Journal of the American Statistical Association* 106 (494): 581–93.

Poole, K. 2007. Changing Minds? Not in Congress! *Public Choice* 131 (3/4): 435–51.

Poole, K., Lewis, J., Lo, J., and Carroll, R. 2011. Scaling Roll Call Votes with Wnominate in R. *Journal of Statistical Software* 42 (14): 1–21.

Rosenbaum, P. R. 2002. *Observational Studies* (2nd ed). New York: Springer-Verlag.

Rosenbaum, P. R. 2007. Interference Between Units in Randomized Experiments. *Journal of the American Statistical Association* 102 (477): 191–200.

Rubin, D. B. 1980. Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Journal of the American Statistical Association* 75 (371): 591–93.

Sobel, M. E. 2006. What Do Randomized Studies of Housing Mobility Demonstrate? *Journal of the American Statistical Association* 101 (476): 1398–1407.

Vote Smart. 2008. *Vote Smart of 2008 New Mexico Key Votes*. Retrieved from http://votesmart.org/bills/NM/2008/ on November 1, 2012.