

Sunshine Works

Comment on “The Adverse Effects of Sunshine:
A Field Experiment on Legislative Transparency
in an Authoritarian Assembly”

James H. Anderson

The World Bank
East Asia and the Pacific Region
Poverty Reduction and Economic Management Department
September 2013



Abstract

Transparency—sunshine—is often touted as a core element of the governance agenda, and one that is most important in environments with low transparency to begin with. In a provocative paper published in the *American Political Science Review*, Edmund Malesky, Paul Schuler, and Anh Tran present the results of a creative experiment in which they provided an additional spotlight on the activities of a random sample of delegates to Vietnam's National Assembly. They report that the effect of sunshine was negative, that delegates subject to this treatment curtailed their speech, and that those who spoke most critically were punished through the subsequent election and promotion processes. The present paper argues that Malesky, Schuler, and Tran's results, if interpreted correctly, actually predict

a net positive effect of transparency. The differences in interpretation stem primarily from three sources: the interpretation of regression results for models with interaction terms, the interpretation of the variable for Internet penetration, and significant pre-treatment differences between treated and control delegates. For the context in which more than 80 percent of delegates operate, Malesky, Schuler, and Tran's results predict a positive but insignificant effect of transparency. In addition, Internet penetration, itself a measure of access to information, is positively associated with critical speech. The paper draws lessons for the design and interpretation of randomized experiments with interaction effects.

This paper is a product of the Poverty Reduction and Economic Management Department, East Asia and the Pacific Region. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at janderson2@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Sunshine Works—Comment on “The Adverse Effects of Sunshine: A Field Experiment on Legislative Transparency in an Authoritarian Assembly”

James H. Anderson*

Keywords: Vietnam; transparency; legislature; political economy; multiplicative interaction; randomized experiments; Internet

JEL Codes: D72, P26, C21, C93

Sector Board: Public Sector Governance (PSM)

*Senior Governance Specialist, World Bank, Hanoi, Vietnam. Email: janderson2@worldbank.org. I thank Edmund Malesky, Paul Schuler, and Anh Tran for supplying their data and programs and for supportive, engaging and constructive discussions during the preparation of this paper. I am grateful to Soren Davidsen, Gabriel Demombynes, Deepak Mishra and Huong Thi Lan Tran for helpful discussions and comments. The support of UK-Aid through the VGEMS trust fund is gratefully acknowledged. The views expressed are my own, and any errors are my responsibility. The findings and interpretations in this paper do not necessarily reflect the views of UK-Aid or of the World Bank, its affiliated institutions, or its Executive Directors.

In *The Adverse Effects of Sunshine: A Field Experiment on Legislative Transparency in an Authoritarian Assembly*, Edmund Malesky, Paul Schuler, and Anh Tran (MST) use a randomized experiment in Vietnam to analyze whether transparency of legislative debates leads to better outcomes. Are legislators (the agents) more responsive to the electorate (the principals) in systems with more transparency? Noting the literature that “legislatures in nondemocratic systems are primarily a forum for contained exchange between the authoritarian leadership and potential opposition”, MST argue that “transparency may have perverse effects”. They summarize their results: “delegates subjected to high treatment intensity demonstrate robust evidence of curtailed participation and damaged reelection prospects. These results make us cautious about the export of transparency without electoral sanctioning.” (p. 762)

MST’s novel experiment was to create personal websites on a popular online newspaper for a randomly selected group of National Assembly (VNA) delegates, providing biographical information and, during the subsequent session of the VNA, a record of their speeches and queries, as well as a “scorecard” on how active, critical, and constituency-minded the delegates’ speeches and queries were. By comparing the activity of the randomly selected delegates before and after this treatment, and to a control group of delegates who were not selected for the treatment, MST sought to tease out the impact of the treatment, which they characterize as “transparency”.

The randomized experiment was clever and useful not only for its intellectual purpose, but also for raising the profile of the VNA. MST should be commended for their innovative approach to seek empirical answers to important questions. As they note, “rolling out initiatives to increase legislative transparency without considering and testing the magnitude of these adverse effects could lead to self-defeating interventions.” (p. 763)

Before proceeding further, a clarification is required, and that is that MST’s experiment was not about “transparency” *per se*, but about putting a spotlight on information that was already publicly available before the experiment. As MST note: “Transcripts of queries and responses are posted on the VNA web site after each session, but these are not presented in an easy-to-find location and have wildly differing titles. Moreover, the transcripts are posted in Word files of two hundred pages or more that make it difficult for citizens to identify quickly what their delegates said.” (p. 770) The experiment, therefore, made information more accessible, but the information was nevertheless accessible to those who were interested and technically capable even before the experiment.¹ This is an important qualifier about the meaning of “transparency” implicit in MST’s study, and will also have implications later in the paper for the proper interpretation of the effect of internet penetration.

In this paper I will argue that transparency’s death has been greatly exaggerated—on balance, MST’s results predict a positive effect of sunshine, not an adverse effect as their title suggests. In Section 1, I will show that the results presented by MST predict that transparency (as defined by them) generally has a positive effect, not an adverse effect, on delegate behavior. In Section 2, I will argue that internet penetration is itself a measure of transparency and that the pattern is as one would expect if there are diminishing returns to information. In Section 3, I will argue that to the extent that there is a significant predicted negative effect of the treatment for delegates in the three or four provinces (out of 64) with the highest levels of internet penetration, this is driven by pre-experiment differences between the treatment and control groups. In Section 4, I will draw attention to the zero-heavy distributions of data for key variables and argue that many of the results, and some misinterpretations, are brought about by these distributions.

¹ Indeed, it was transparency prior to and after the experiment that permitted MST to construct their dataset.

In Section 5, I will argue that MST's analysis of reward and punishment is less robust than it appears. In Section 6, I will summarize and offer some closing thoughts on the methodological lessons for research using randomized experiments and interaction terms, as well as the need to differentiate between theories of debate, criticism, and authoritarianism, on the one hand, and transparency's effect on debate, on the other.

1. *Interpretation of interaction terms*

A key set of results in MST surround the impact of the transparency experiment on delegate behavior during the 6th session of the VNA, focusing in particular on the number of *questions* asked, the percentage of *critical* questions, and the number of *speeches* made. The first two are understood to be more challenging and responsive to constituents, while the latter is understood to be consistent with conformist behavior. MST use a difference in differences approach, such that the dependent variable is either the increase in activity between the 5th and 6th session, or the increase between the average of sessions 1-5 and the 6th session. On the right hand side of their regressions, MST include a number of delegate characteristics, characteristics of the provinces they represent, and two key variables of interest and their interaction term. The variables of interest are a dummy for *treated* delegates, the level of *internet* penetration in the province, and the interaction between the two (*treated*internet*). In a table on page 777 of their paper, MST present the results of ten regressions and on page 778 they graph the predicted marginal impact of *treatment* for various levels of *internet* for four of those regressions, the ones controlling for delegate and province level characteristics. The four key sets of regression results are reproduced below in columns (1) – (3) of Table 1. The coefficients on the delegate and province characteristics are omitted as they do not depend on *treatment* or *internet* penetration and therefore are not pertinent to this discussion.

MST's first finding is that the direct effect of *treatment* is insignificant indicating that “the transparency intervention did not lead to improved delegate performance or activity.” (p. 773) This finding, clearly, did not lead to the strong conclusions embodied in the title of the paper, the “adverse effects of sunshine.” That finding comes, rather, comes from a set of regressions that go beyond the simple comparisons of *control* and *treatment* groups. MST summarize the most striking feature of their results which relates to the coefficient on the interaction term, *treated * internet*.

Internet penetration significantly magnifies the impact of the treatment in a negative direction for the number of questions asked and the percentage of critical queries. These results appear to be robust across specifications. Substantively, each additional Internet subscription per 100 citizens is associated with a 0.18 reduction in the number of questions asked in the treated group and a 1.9% decrease in the percentage of critical queries between the fifth and sixth sessions. Thus, when Internet penetration is about 8% (the level observed in Hanoi and Ho Chi Minh City), we find that treated delegates ask a full question less and reduce their criticism more than 12% below the delegates in the control group—a highly significant difference, as measured by the t-value over 6. When we compare the treatment intensity between the sixth session and average participation in Models 9 and 10, we see similar though slightly less pronounced results. Here, the effect is a reduction of about 0.6 questions and 0.8% less criticism in the fully specified model. (776-778)

As both *internet* and *treatment* are interacted with each other, the coefficients need to be examined in tandem rather than independently.² In all cases, the intercept is positive and the slope is negative, meaning that for some levels of *internet* penetration the predicted effect of *treatment* is positive, and for others negative. The net effect of *treatment*, therefore, is ambiguous and depends on the level of *internet* penetration. Column (4) shows the critical value of *internet* for which the effect of *treatment* is predicted to have zero effect.

The charts that MST produce showing the effect of *treatment* conditional on level of *internet* penetration correspond to the equations in Table 1 and follow exactly the approach described above. While the charts may accurately depict the intercept and slopes of the lines, however, they are misleading. The charts do not depict accurately the underlying distribution of the variable *internet*. Although the charts presented by MST seem to imply a uniform distribution for *internet*, the distribution of *internet* is, in fact, highly skewed. As MST reported in the article, *internet*, which is measured at the province level, ranges between 0.22 and 8.63, with a mean of 1.39 for the treatment group and 1.28 for the control group. Figure 1 depicts a simple histogram of the variable across provinces (left panel) and across delegates (right panel). As an example of how the distribution affects our interpretation of the results, the figure also depicts the critical value of *internet* at which the effect of *treatment* on the change in *question* count between session 5 and 6 ($Q_{5,6}$) is zero.

As Figure 1 makes clear, the vast majority of provinces (and therefore delegates) have levels of *internet* penetration that are below the critical value where the predicted impact of *treatment* moves from positive to negative. Column (5) of Table 1 shows the percentage of delegates representing provinces for which the model predicts a positive impact of *treatment*, i.e., a positive predicted effect of transparency. For three out of the four regressions, the model actually predicts that the transparency experiment increased delegate activity for 80-84% of delegates. Thus, while MST suggest that the findings support the hypothesis that transparency led delegates to curtail their speech, their results largely suggest the opposite, a finding that is inconsistent with the “adverse effects of sunshine.”

Figure 2 makes this point graphically, reproducing the charts as reported by MST. Below each of MST’s charts I have created the same chart, except that it more accurately represents the distribution of *internet* across provinces. Each mark represents a province. (A similar chart capturing the distribution of delegates would look identical since *internet* penetration varies only by province.) It bears emphasizing that the data for *internet* covers virtually the entire population, rather than a sample, and that it is discrete, not continuous.

The charts produced by MST are different from the reproduction in Figure 2 in one respect, and that is that MST’s charts included range bars representing 90% confidence intervals. These have been omitted from Figure 2 for presentation purposes, but the issue of statistical significance merits further discussion.

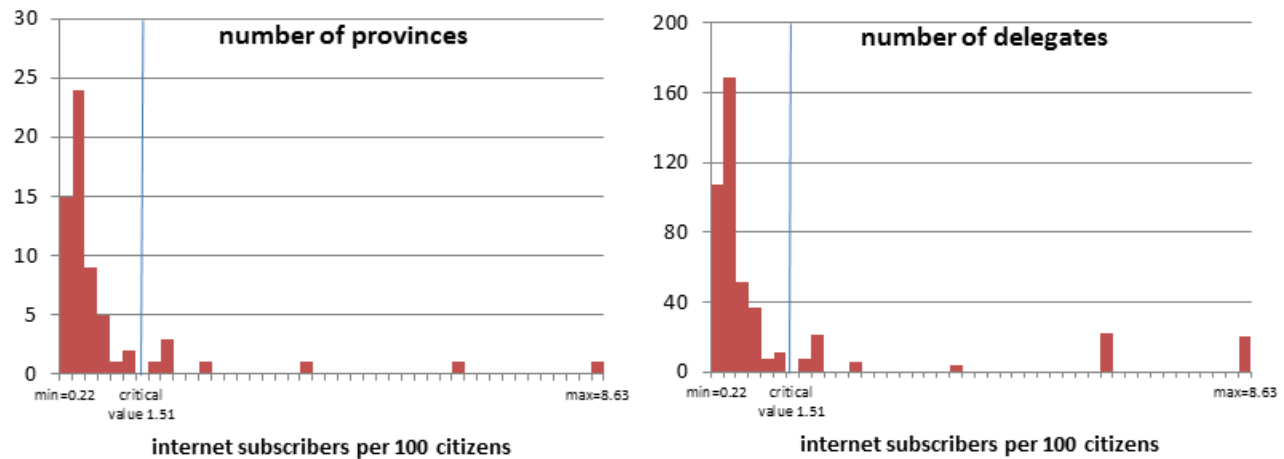
² See Braumoeller (2004) and Brambor, Clark, and Golder (2006) for useful discussions of multiplicative interaction terms.

Table 1. Predicted impact of treatment on challenging delegate behavior, conditional on level of internet penetration

Dependent variable (description)	Dependent variable (shorthand)	<i>treated</i> (1)	<i>internet</i> (2)	<i>treated</i> * <i>internet</i> (3)	Critical value (level of <i>internet</i> at which treatment has zero effect) (4)	% of delegates in the internet range for which the model predicts a <i>positive</i> effect of treatment (5)	% of delegates in the internet range for which the model predicts a <i>negative</i> effect of treatment (6)
Increase in <i>Question</i> count (#) between session 5 and session 6	$Q_{6,5} = Q_6 - Q_5$	0.271* (0.154)	0.105 (0.070)	-0.179*** (0.027)	1.51	83%	17%
Increase in <i>Critical</i> questions (%) between session 5 and session 6	$C_{6,5} = C_6 - C_5$	2.799 (1.907)	2.850*** (0.843)	-1.865*** (0.330)	1.50	83%	17%
Increase in <i>Question</i> count (#) between the average of sessions 1-5 and session 6	$Q_{6,1-5} = Q_6 - Q_{avg1-5}$	0.021 (0.129)	0.048 (0.033)	-0.058** (0.027)	0.36	18%	82%
Increase in <i>Critical</i> questions (%) between average of sessions 1-5 and session 6	$C_{6,1-5} = C_6 - C_{avg1-5}$	0.937 (1.762)	2.247*** (0.319)	-0.738*** (0.276)	1.27	80%	20%

Columns (1) – (3) partially reproduces equations 4, 8, 9, and 10 from MST Table 5. The coefficients for other variables that were included in the regressions are not reproduced here as they drop out when taking derivatives with respect to *treated* or *internet*. The numbers in parentheses are the robust standard errors, clustered at the provincial level. *** p<0.01, ** p<0.05, * p<0.1. Column (4) is determined by taking the derivative (or difference) of the fitted equations with respect to *treatment* and setting equal to zero. Columns (5) and (6) depicts the proportion of delegates in provinces with *internet* penetration below and above that threshold, respectively.

Figure 1. Distribution of the variable *internet*



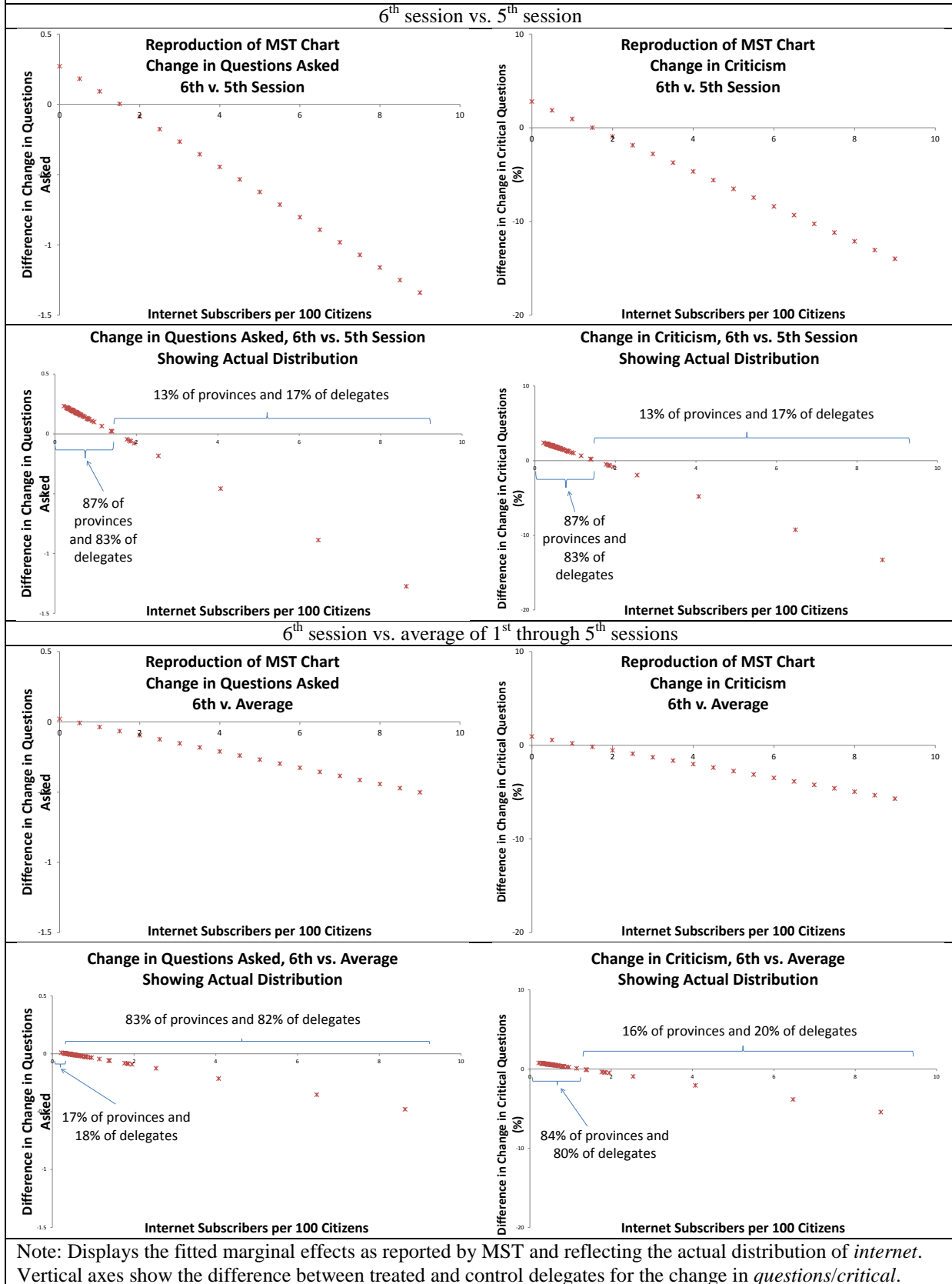
Notes: Figures show histograms of the variable *internet*, across provinces (left) and across delegates (right). The “critical value” is the level of *internet* penetration at which MST’s regressions predict that the effect of *treatment* moves from positive to negative.

According to MST, the predicted effect of *treatment* is not statistically significant over the range of *internet* at which the predicted effect of *treatment* is positive. This is true. It is also true, however, for much of the range over which the predicted effect of *treatment* is negative. For example, for the change in *critical* questions between the 5th and 6th sessions, there are only three provinces (Hanoi, HCMC, and Da Nang) out of 64 (i.e., fewer than 5 percent of provinces) for which the predicted effect of *treatment* is statistically significantly different from zero at any conventional level.³

At the same time, this hardly seems a satisfying approach to identifying the effect of a treatment in an experiment. What we primarily want to know is simple: on average, did the treatment have an effect, and if so in what direction? As noted earlier, the direct effect of *treatment* provides the first answer to this question, and this should be fairly compelling. After all, the attraction of a truly randomized experiment with well selected samples is that one should not need to control for much else. Based on MST’s direct effect analysis, the *treatment* “did not lead to improved delegate performance or activity,” but neither did they lead to worse delegate performance or activity—the reported coefficients were all positive and insignificant. What of the regressions that included the interaction terms, the regressions that demonstrate the adverse effects of sunshine? For the median value of *internet*, three out of the four key regressions predict a positive (but insignificant) effect of transparency, and the same is true at the mean of *internet* and one standard deviation above and below.

³ It should be noted that one of those three provinces, Da Nang, did not have any delegates in the *treatment* group. There were another five provinces with no delegates in the treatment group: Can Tho, TT-Hue, Dien Bien, Vinh Phuc, and Hoa Binh. Da Nang and Can Tho are two of the five large centrally managed cities in Vietnam.

Figure 2. Intensity of Treatment Effect as presented by MST and showing actual distributions



More directly, one would ultimately like to know the aggregate effect of the *treatment*, and since the *treatment* and *control* groups combined make up virtually the entire population of VNA delegates, this aggregate effect is of particular interest. Over the 462 delegates in the experiment, after controlling for a range of delegate and province level characteristics and controlling for the level of *internet* penetration in the province, what was the aggregate effect of the *treatment*? For three out of the four key specifications, the sum of the predicted change in delegate behavior was positive, not negative. In other words, the very large number of small positive effects for delegates representing provinces with low *internet* penetration outweighs the small number of large negative effects in provinces with high *internet* penetration. Again, the balance of evidence predicts, if anything, a positive effect of transparency, not the reverse. While the statistical (in)significance of these predicted positive effects does not obviate the possibility of no effect, the sign of the predicted aggregate effect is sufficient to cast doubt on the aggregate “adverse effect of sunshine.”

2. *Interpretation of Internet penetration*

While the primary interest is the aggregate effect of transparency, a second question is worth pursuing. In what situations does transparency have a positive or negative effect? Here, MST’s finding that the effect of transparency is more negative for delegates representing provinces with high levels of *internet* penetration, even if only statistically significant for 5 or 6 percent of provinces, deserves discussion. Why might this be so? In this section and the next I present two alternative explanations, neither of which excludes the other.

The variable *internet* is the lynchpin for MST’s argument that more transparency led to curtailed criticism by deputies—regressions without the interaction term for *internet* penetration showed no direct effect of the *treatment*. Getting the interpretation of “*internet* penetration” correct, therefore, is essential.

MST interpret *internet* penetration as measuring the “intensity of treatment” since the treatment was the publishing of web-based pages for *treated* delegates. However, there is an alternative and simpler interpretation of the variable. Rather than interpreting *internet* penetration as a measure of “treatment intensity,” as MST do, one must recognize that *internet* penetration *is itself a measure of access to information and to transparency*. As noted earlier, MST’s experiment did not really make transparent something that was previously hidden; rather their experiment put a spotlight on information that was already available on the internet. Even beyond the transcripts of VNA sessions, it is clear that people living in provinces with higher levels of *internet* penetration have more information at their disposal about both local and national level issues, including discussions at the VNA, than people living in places with low levels of *internet* penetration. If *internet* penetration is itself a measure of transparency, then the patterns reported by MST make perfect sense: the positive effect of transparency is highest in places where people were most starved of information in the first place, those with low levels of *internet* penetration. Indeed, as we will see below, after the experiment, for both *control* and *treatment* groups, *internet* penetration remained positively correlated with both the number of *questions* asked by delegates and the percentage of *critical* questions for many specifications. Even in the difference in differences analysis that MST presented, *internet* is positively related to the proportion of *critical* questions for both *control* and *treatment* groups, and for the number of *questions* for the *control* group.

3. *Pre-treatment differences between treatment and control groups*

Section 1 took MST's results from the difference in differences analysis at face value and argued that the results predict a net positive (but insignificant) effect of the *treatment*. As both the direct effect of the treatment and the aggregate effect of the treatment in specifications with the interaction term were positive and insignificant, those findings are sufficient to cast doubt on the "adverse effect of sunshine", but not sufficient to suggest that "sunshine works." Section 2 also took MST's results at face value but argued that *internet* should be interpreted as transparency and, with that interpretation, that the difference in differences analysis shows a mostly positive effect of transparency. In the next two sections, I continue to take MST's specification at face value but look more closely at the randomized experiment and the difference in differences analysis.

Prior to any randomized experiment one should examine key relationships to ensure that the *treatment* and *control* groups are sufficiently similar that the difference in outcome can be attributed to the *treatment*.⁴ As the relationship between delegate activity and *internet* penetration is central to MST's conclusions, some attention to their relationship prior to the experiment is due.

What was the relationship between the variables for delegate activity and *internet* before and after the experiment? For session 5, before the experiment, there is a strong positive relationship between *internet* and *questions* for those who were subsequently "treated", but a negative relationship for those in the control group. (Table 2) The same pattern holds when examining the relationship between *internet* and the percentage of *critical* questions. The final four rows of Table 2 show that these differences are highly statistically significant, with different intercepts and slopes for the *treatment* and *control* groups. As this was prior to the *treatment*, this pattern must be a matter of happenstance, but it nevertheless suggests a problem with MST's difference in differences approach using *internet* as an interaction variable.⁵

The top panel of Figure 3 makes this point graphically, showing the relationship between *internet* penetration and delegate behavior during the 5th session of the VNA, prior to the experiment. Even before the experiment, there were stark differences between the patterns of the *control* and *treatment* groups. A simple process of mean reversion moving from session 5 to session 6 would have generated exactly the patterns that MST observe: In places with high *internet* penetration, the difference between sessions 5 and 6 would be strongly positive for the *control* group and strongly negative for the *treatment* group. The reverse pattern would obtain for provinces with low levels of *internet* penetration.

Indeed, this is exactly what we observe. During session 6, the pattern between *questions* and *critical* on the one hand, and *internet* on the other, is more similar between *control* and *treatment* groups than was the case during session 5 (Table 3). These relationships are also

⁴ See, for example, the related discussion on cross-cutting designs, evaluating two treatments simultaneously, and the need for baseline surveys in Duflo, Glenmeister and Kremer (2006).

⁵ As Abadie (2005) wrote "it is well known that the DID estimator is based on strong identifying assumptions. In particular, the conventional DID estimator requires that, in the absence of the treatment, the average outcomes for the treated and control groups would have followed parallel paths over time. This assumption may be implausible if pre-treatment characteristics that are thought to be associated with the dynamics of the outcome variable are unbalanced between the treated and the untreated." (p. 1)

Table 2. Pre-“treatment” relationship between challenging delegate behavior and internet penetration

Dependent variable	Group of delegates	<i>constant</i>	<i>internet</i>	<i>treated</i>	<i>treated</i> <i>*internet</i>	other variables	R ²	N
(1) Q ₅ . <i>Question</i> count (#) in session 5	“control” group of delegates	0.370*** (0.075)	-0.041*** (0.012)				0.005	318
(2) Q ₅	“treatment” group of delegates	0.116 (0.087)	0.176*** (0.025)				0.092	144
(3) Q ₅	all delegates	0.287*** (0.063)	0.034* (0.017)				0.004	462
(4) Q ₅	“control” group of delegates	0.904 (0.837)	-0.058 (0.043)			Included	0.043	318
(5) Q ₅	“treatment” group of delegates	0.981 (1.542)	0.211** (0.102)			Included	0.132	144
(6) Q ₅	all delegates	1.172 (0.719)	0.054 (0.046)			Included	0.032	462
(7) C ₅ . <i>Critical</i> questions (%) in session 5	“control” group of delegates	2.762*** (0.812)	-0.298** (0.122)				0.003	318
(8) C ₅	“treatment” group of delegates	-0.074 (0.966)	1.756** (0.370)				0.098	144
(9) C ₅	all delegates	1.841*** (0.633)	0.408* (0.229)				0.005	462
(10) C ₅	“control” group of delegates	13.243* (7.476)	-1.377 (0.988)			Included	0.037	318
(11) C ₅	“treatment” group of delegates	12.938 (11.856)	2.933*** (0.607)			Included	0.174	144
(12) C ₅	all delegates	15.045** (6.442)	-0.017 (0.961)			Included	0.025	462
(13) Q ₅ . <i>Question</i> count (#) in session 5	all delegates	0.370*** (0.075)	-0.041*** (0.012)	-0.253** (0.102)	0.217*** (0.027)		0.037	462
(14) C ₅ . <i>Critical</i> questions (%) in session 5	all delegates	2.762*** (0.812)	-0.298** (0.122)	-2.837** (1.293)	2.053*** (0.373)		0.033	462
(15) Q ₅ . <i>Question</i> count (#) in session 5	all delegates	1.240* (0.731)	-0.036 (0.042)	-0.298*** (0.107)	0.223*** (0.027)	Included	0.067	462
(16) C ₅ . <i>Critical</i> questions (%) in session 5	all delegates	15.812** (6.374)	-0.864 (0.943)	-2.937** (1.364)	2.112*** (0.418)	Included	0.054	462

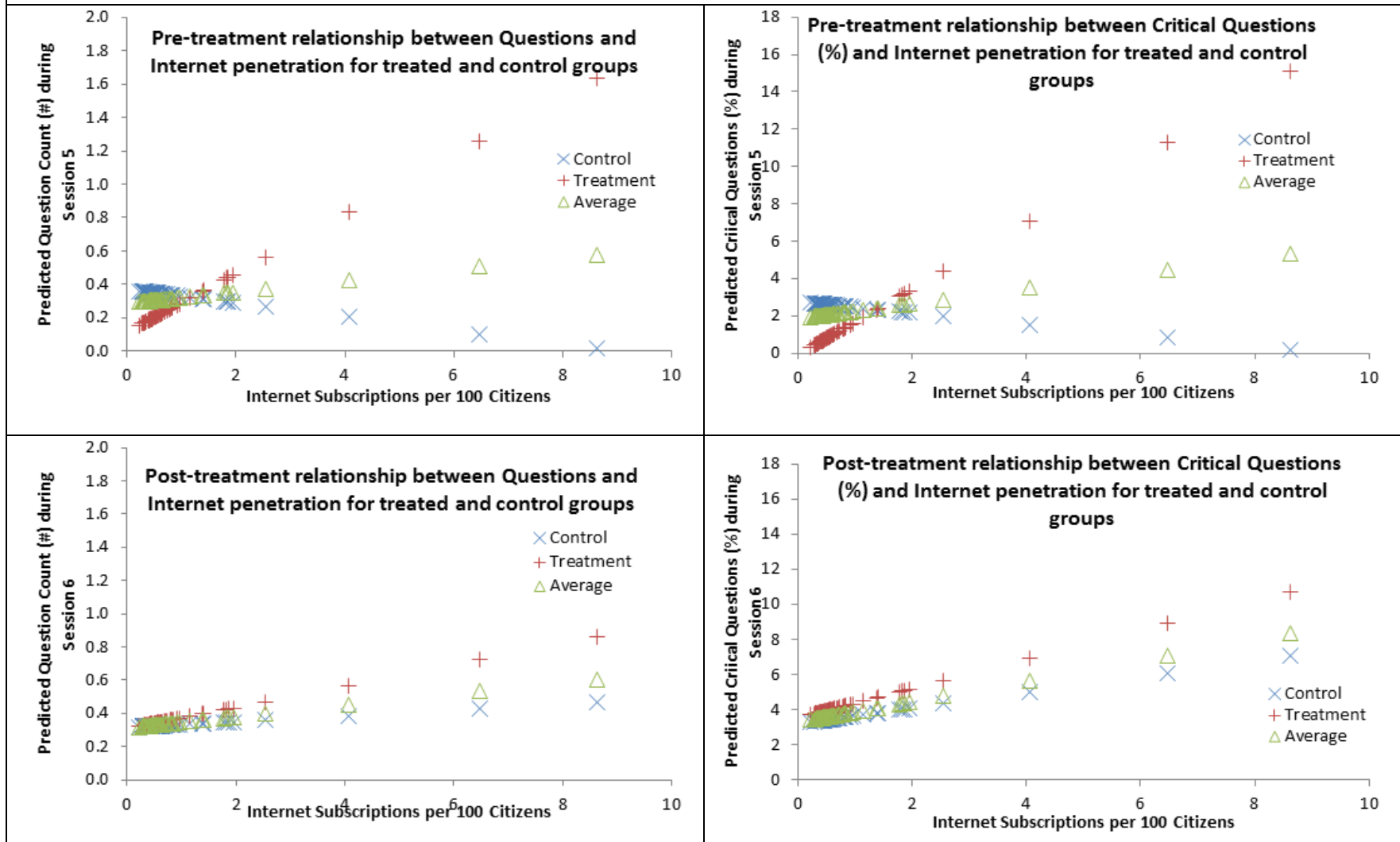
Simple LS regressions. Following MST, the numbers in parentheses are the robust standard errors, clustered at the provincial level. The additional explanatory variables are the same as those in MST. *** p<0.01, ** p<0.05, * p<0.1.

Table 3. Post-“treatment” relationship between challenging delegate behavior and internet penetration

Dependent variable	Dependent variable (shorthand)	<i>constant</i>	<i>internet</i>	other right-hand side variables	R ²	N
(1) Q ₆ - <i>Question</i> count (#) in session 6	“control” group of delegates	0.314*** (0.082)	0.017 (0.033)		0.000	318
(2) Q ₆	“treatment” group of delegates	0.306** (0.127)	0.064* (0.034)		0.011	144
(3) Q ₆	all delegates	0.311*** (0.064)	0.034 (0.034)		0.003	462
(4) Q ₆	“control” group of delegates	0.488 (0.741)	0.041 (0.065)	Included	0.039	318
(5) Q ₆	“treatment” group of delegates	-0.151 (1.344)	0.175** (0.075)	Included	0.079	144
(6) Q ₆	all delegates	0.420 (0.543)	0.088* (0.046)	Included	0.039	462
(7) C ₆ - <i>Critical</i> questions (%) in session 6	“control” group of delegates	3.201*** (0.979)	0.446 (0.603)		0.004	318
(8) C ₆	“treatment” group of delegates	3.526* (1.840)	0.829 (0.714)		0.010	144
(9) C ₆	all delegates	3.289*** (0.869)	0.583 (0.641)		0.006	462
(10) C ₆	“control” group of delegates	19.465* (11.434)	1.890*** (0.480)	Included	0.077	318
(11) C ₆	“treatment” group of delegates	11.379 (23.050)	2.205** (1.087)	Included	0.048	144
(12) C ₆	all delegates	17.531* (9.250)	2.094*** (0.459)	Included	0.060	462

Notes: Simple LS regressions. Following MST, the numbers in parentheses are the robust standard errors, clustered at the provincial level. The additional explanatory variables are the same as those in MST. *** p<0.01, ** p<0.05, * p<0.1.

Figure 3. Pre-treatment and Post-treatment relationship between delegate behavior and internet penetration



Note: Top panels plot the pre-treatment fits between *internet* and *questions* (equations (1)-(3) in Table 2 (top left)) and *critical* questions (equations (7)-(9) in Table 2 (top right)). Bottom panels plot the post-treatment fits between *internet* and *questions* (equations (1)-(3) in Table 3 (left)) and *critical* questions (equations (7)-(9) in Table 3). All panels plot over the actual range of the variable *internet*.

depicted graphically in the bottom panel of Figure 3. MST's key finding that the *treated* delegates in areas with high *internet* penetration reduced their *questions* and percentage of *critical* questions vis-à-vis the *control* group reflect the differences between the groups before the experiment, not after.

4. *A remark on the distribution of key variables*

If the reader finds the patterns reported in these tables and charts odd, it is because they all depict average relations between variables whose distributions are highly skewed or have many zeros. We have already seen that *internet* has a highly skewed distribution with a long tail, and this influences the interpretation of the results. In fact, however, it is also instructive to examine the distributions of the key dependent variables. In both sessions 5 and 6, more than 90% of delegates did not ask even one *question*. In sessions 5 and 6, more than 95% and 92% of delegates, respectively, had zero *critical* questions.⁶

The implication of these zero-heavy distributions is that the difference variables, which serve as the dependent variables in MST's key regressions, are also mostly zeros: 86% in the case of $Q_{6,5}$ and 91% in the case of $C_{6,5}$. The distributions are provided in Table 4. Such distributions may raise questions about the choice of regression techniques—MST use OLS for the regressions with $Q_{6,5}$ and $C_{6,5}$ as the dependent variables—but also raise the risk that a small number of observations can have important effects on results. As we will see, this does seem to be the case for the key regressions relating delegate behavior, transparency, and post-treatment election and promotion prospects.

5. *Impact of transparency on delegates' reelection and promotion*

In addition to the finding that transparency leads to curtailed critical speech, MST also explored hypotheses surrounding the impact of such speech, running 36 regressions relating *speeches* and *questions* to proxies for electoral rewards (such as probability of renomination and share of vote obtained), for elite punishment and reward (such as being put in difficult-to-win districts or being promoted), and for voter responsiveness. The key right-hand side variables (*speeches* and *questions* in session 6) were interacted with the *treatment* dummy for 24 of the regressions, while for the other 12 regressions the *treatment* dummy was alone on the right-hand side. (*Internet* and its interaction with *treatment* were omitted from these 36 regressions.)

⁶ The patterns in which such a large proportion of delegates did not ask questions is not merely one of random factors, nor of choice by delegates themselves. Time is rationed in the VNA and delegates who wish to speak are not necessarily given an opportunity to do so, even if they so desire. (World Bank and others, 2009) This raises a methodological concern about whether the dependent variables could capture the effect of treatment, and for how cleanly one may expect treatment to influence the outcome. MST's reference to one third of delegates speaking and criticizing is much larger than the 10% I am reporting in this paper. The 10% figure is correct.

Table 4. Distribution of key dependent variables in MST's results

Change in <i>Questions</i>			Change in Percentage of <i>Critical Questions</i>		
$Q_{6,5} = Q_6 - Q_5$	N	percentage	$C_{6,5} = C_6 - C_5$	N	percentage
-7	1	0.2%	-91 to -100	1	0.2%
-6	1	0.2%	-81 to -90	0	0.0%
-5	5	1.1%	-71 to -80	1	0.2%
-4	4	0.9%	-61 to -70	1	0.2%
-3	3	0.7%	-51 to -60	0	0.0%
-2	12	2.6%	-41 to -50	3	0.7%
-1	5	1.1%	-31 to -40	2	0.4%
0	398	86.2%	-21 to -30	3	0.7%
1	6	1.3%	-11 to -20	1	0.2%
2	10	2.2%	-1 to -10	1	0.2%
3	8	1.7%	exactly 0	422	91.3%
4	2	0.4%	+1 to +10	2	0.4%
5	4	0.9%	+11 to +20	2	0.4%
6	1	0.2%	+21 to +30	3	0.7%
7	1	0.2%	+31 to +40	6	1.3%
8	0	0.0%	+41 to +50	5	1.1%
9	0	0.0%	+51 to +60	1	0.2%
10	0	0.0%	+61 to +70	3	0.7%
11	0	0.0%	+71 to +80	1	0.2%
12	1	0.2%	+81 to +90	0	0.0%
Total	462	100%	+91 to +100	4	0.9%
			Total	462	100%

Note: The table is restricted to the same delegates as in the experiment.

Out of the 12 regressions with only the *treatment* dummy on the right-hand side, three had significant coefficients. At the 5% level of significance, MST report that *treated* delegates were more likely to be placed in easier to win districts, as measured by the number of *seats per candidate*. At the 10% level of significance, the regressions suggest that the *official turnout* was lower in districts with *treated* delegates, and that *treated* delegates were less likely to be *renominated*. Although *treated* delegates were more likely to be placed in easy to win districts, the focus is on the chance of reelection:

In sum, the general equilibrium effect of the transparency treatment is strongly negative. Treated delegates were 9.5% less likely than control delegates to be renominated for seats. Treated delegates who were renominated were 4.6% less likely to be reelected and retain their seats. The probability of election is not significant, because the baseline probability of reelection for renominated incumbents is 92%. The net effect, however, is that treated delegates were about 10% less likely to retain office than their peers in the control group, a statistically significant finding. Despite the fact that many treated delegates actually curtailed their behavior from previous sessions and most delegates increased their visible effort, these tactics were not enough. Their

enhanced visibility was still threatening enough for regime and provincial leaders to keep them out of office. (p. 780)

The latter finding that *treated* delegates were about 10% less likely to retain office is less impressive, however, when one considers that this only relates to the treatment, and takes no account for whether or not the delegate actually spoke. As we have seen, some 90% of the delegates did not ask any *questions*. In fact, when parsing the data according to whether or not the delegate asked any *questions*, it is clear that the smaller renomination probability for *treated* delegates was driven by those who did not ask *questions*, rather than by those who did (Table 5). If one follows the logic that renomination was influenced by asking questions in the spotlight, these results suggest that, if anything, those who availed themselves of the visibility of the transparency experiment by asking questions were rewarded and those who remained quiet were punished, the precise opposite of MST’s conclusion.

Table 5. Probability of renomination for control and treated groups, by those who asked questions and those who did not

	(1) All delegates		(2) Delegates who DID NOT ask questions in session 6		(3) Delegates who DID ask questions in session 6	
	Control	Treated	Control	Treated	Control	Treated
Delegate not renominated	196 (62%)	100 (70%)	180 (62%)	93 (73%)	16 (57%)	7 (44%)
Delegate renominated	122 (38%)	43 (30%)	110 (38%)	34 (27%)	12 (43%)	9 (56%)
Column totals	318 (100%)	143 (100%)	290 (100%)	127 (100%)	28 (100%)	16 (100%)

After first suggesting that *treated* delegates were punished for being in the treatment group, MST go on to acknowledge that the regressions with only the *treatment* dummy on the right-hand side miss the point, since some treated delegates may behave in a conformist manner, obviating the need for any form of electoral or elite punishment. The real question is whether those who were treated *and* spoke out in a critical way were subject to some form of punishment. They address this question by interacting the *treatment* dummy with the number of debate *questions* and the number of *speeches*. The 12 regressions with debate *questions* as the variable of interest are the key ones since they address the hypotheses about the post-treatment impacts on delegates who spoke up in the sunshine. Of those 12 regressions, only three had results worth remarking upon.⁷

The first of these three sets of findings surrounds the delegates’ post-treatment reelection prospects. MST note “very little evidence that post-treatment behavior affected the electoral results, with one notable exception—vote share.”

⁷ As this part of MST’s research focused on how delegates who were subjected to the *treatment* and spoke out were rewarded or punished after the fact, the locus of attention is correctly placed on the interaction term, not on the overall effect of the *treatment*.

Here, we find that the interaction between treatment and question-asking is significant and negative, whereas the component term on treatment is insignificant (albeit positive). This implies that delegates who did not curtail their sensitive questions and criticisms in the presence of transparency received significantly smaller shares of votes than their silent peers (about 6% less for each question asked). (p. 783)

MST's finding that outspoken delegates in the *treated* group received significantly smaller *vote shares* than their quieter peers appears, however, to be the result of a programming error. Rather than *vote share*, the dependent variable which generated the results MST reported was a different (binary) variable entirely. When using *vote share* as the dependent variable, the results disappear from significance altogether (Table 6).

The second of the three sets of noteworthy results centers on hypotheses about rewarding and punishing delegates for their speech.

[W]e find compelling evidence of leaders punishing and rewarding delegates for upholding the co-optive bargain. Treated delegates tended to be placed in easier-to-win districts, as measured by the seats-to-candidates ratio. Calculating the substantive effects of the ordered probit model reveals that treated delegates were 6% less likely to be placed in districts with a 50% probability of victory and 12% more likely to be placed in districts with a 67% probability of victory. This provides tentative (and highly speculative) evidence for H3 that central officials tried to buy off complicity of treated delegates by offering greater opportunities for legislative victory. Nevertheless, in line with H6, when this bargain was not upheld and treated delegates spoke up during the session, they found themselves saddled with more challenging electoral placements. For each question asked during the query session, a delegate had a 3% higher probability of being placed in one of the most difficult-to-win districts, where there were twice as many candidates as seats. These same delegates were 5% less likely to be placed in one of the easier-to-win (67% probability) districts. Interestingly, nontreated delegates who continued to speak were actually rewarded slightly, with each question gaining them a 2% higher probability of being placed in an easy district.

As noted earlier, however, the zero-heavy distributions depicted in Table 4 leads to the risk that a small number of observations can exert heavy influence. Indeed, a single observation has a strong influence on these results. The most active delegate in the experiment asked 14 questions in the 6th session of the VNA, and this delegate was in the control group. When this single delegate is excluded from the analysis, the finding that treated delegates who asked questions were placed in electoral districts with higher *seat per candidate* disappears from significance entirely. I emphasize that my point is not that such influential observations should be excluded; they are part of the distribution and should not be excluded *a priori*. Rather, my point is that when variables have heavily skewed distributions, a small number of observations can have important effects on the results, and modesty should accompany representations of the robustness of those results (see Table 6).

The third set of noteworthy results surrounded the idea that *promotions* are used as a reward incentive:

Moving beyond the electoral domain to further probe the punishment/reward hypotheses, we find further evidence of the use of promotion to leadership positions in the VNA or ministries as a reward incentive. The coefficient on treated delegates is positive, but nonsignificant, so there is only the slimmest of evidence for preemptive reward. Nevertheless, for each question a delegate asked, the marginal probability of promotion to higher office declined by nearly 5% from the baseline probability of 22%. Again, this provides tentative evidence for H3—when delegates did not uphold the co-optive bargain, central officials chose to punish their transgressions. (783-784)

This conclusion is based on a set of regressions in MST’s Table 7 that find a positive and insignificant coefficient on both *treated* and *questions*, but a weakly significant (10% level) negative coefficient on the interaction term. Although the table’s notes indicate the results were generated from a probit regression, in fact the results that are presented were drawn from an OLS regression, despite the fact that the dependent variable was binary. Attempts to use probit, however, failed as the interaction term is always associated with no *promotion*. Again, the issue of the zero-heavy distribution makes the results quite sensitive. A look at the distribution of the data is telling. Table 7 shows a cross-tab of the number of *questions* asked in session 6, for *treated* and *control* groups, and for *promoted* and *non-promoted*, delegates. The table is restricted to the same 164 delegates used by MST in analyzing the possibility that *promotions* were offered to those who did not speak out in the spotlight.

Table 6. Correction for programming error and examining the impact of influential observations on electoral results, placement in easy-to-win districts, and promotions

Dependent variable (description)	<i>constant</i>	<i>treated</i>	<i>questions session 6</i>	<i>treated * questions session 6</i>	(pseudo) R ²	N
1 <i>Vote share</i> (reproduced from MST Table 7, 1.3 (3)) in which a coding error led to the use of the wrong dependent variable	0.094*** (0.034)	0.039 (0.045)	0.048** (0.017)	-0.060*** (0.020)	0.763	165
2 <i>Vote share</i> (matching MST Table 7, 1.3 (3)), but using the correct dependent variable	72.8*** (2.26)	-0.829 (2.10)	-0.551 (0.48)	-0.854 (1.02)	0.048	165
3 <i>Seats per candidate</i> (reproduced from MST Table 7, 2.1 (3))	-1.114*** (0.191)	0.649** (0.256)	0.092* (0.055)	-0.213** (0.101)	0.0476	164
4 <i>Seats per candidate</i> (matching MST Table 7, 2.1 (3) but dropping a single observation)	-1.129*** (0.191)	0.630** (0.256)	0.025 (0.094)	-0.146 (0.127)	0.0405	163
5 <i>Promoted</i> (reproduced from MST Table 7, 2.3 (3))	0.276* (0.143)	0.019 (0.063)	0.006 (0.024)	-0.047* (0.026)	0.128	164
6 <i>Promoted</i> (matching MST Table 7, 2.3 (3) but dropping a single observation.)	0.293* (0.142)	0.018 (0.063)	-0.011 (0.017)	-0.030 (0.021)	0.123	163

This table partially reproduces equations from MST Table 7. The coefficients for other variables that were included are not reproduced here as they are not pertinent and as they were not presented in MST Table 7. These include central nominated, full-time, and either retired or age. (MST Table 7 used retired for some and age for others. The present regressions mimic MST’s selections exactly.) The numbers in parentheses are the robust standard errors, clustered at the provincial level. *** p<0.01, ** p<0.05, * p<0.1. Equation 1 uses the same method and data as MST, however a programming error meant that the wrong dependent variable was used. Equation 2 uses the same right hand side variables as equation 1 for comparability, but uses the correct variable for “vote share” as the dependent variable. Equations 3 and 4 use ordered probit, following MST. Equations 5 and 6 use OLS as that is the procedure used by MST, despite the binary dependent variable. Attempts to use probit fail because the interaction term always maps to *promoted*=0.

As Table 7 makes clear, the data underlying this analysis do not tell a convincing story about the links between *questions*, *promotion* and *treatment*. Of the 21 delegates who were

promoted, only two asked any *questions* at all, and both were in the *control* group.⁸ If either one of these two were not included in the regression, the significance of the interaction term in the (mis-specified) OLS regression falls below conventional levels (Table 6). As noted above, this does not mean that such influential observations should be excluded *a priori*, just that the fragility of results should be disclosed. The narrative that “for each question a delegate asked, the marginal probability of promotion to higher office declined by nearly 5%” is inconsistent with a distribution in which zero treated, and only two control, delegates asked questions and were promoted.

Table 7. Distribution of delegates by question, treatment, and promotion

Q ₆	Control		Treated		Total
	Not Promoted	Promoted	Not Promoted	Promoted	
0	93	13	31	6	143
1	2		1		3
2	3		3		6
3	1		2		3
4	1		1		2
5	1				1
6	1	1	1		3
8		1	1		2
14	1				1
Total	103	15	40	6	164

Note: Table is restricted to the same 164 observations in MST Table 7, Panel 2.3 (3), i.e., those with full information for promotion, questions, treatment, centrally nominated, age and fulltime.

In summary, while MST’s attempts to examine the impact of the spotlight treatment on the delegates who spoke up, those who “did not uphold the co-optive bargain,” is interesting, the results are not compelling enough to warrant the conclusions. Only three of the 12 key regressions in MST’s Table 7 had noteworthy results: one was due to a programming error, and the other two are highly sensitive to even a single observation. The fragility of these results should add additional pause to our confidence in the strength and robustness of the “adverse effects of sunshine”.

6. *Closing thoughts*

The experiment carried out by MST was both innovative and important. It was innovative in its attempt to introduce transparency to deliberative processes in a random way, facilitating statistical analysis of the impact of that form of transparency. It was important in that it also provided a valuable service for both Vietnamese citizens, who got to know their

⁸ It should be remembered that MST’s experiment excluded all Politburo members, as well as five high-profile delegates who were given webpages but excluded from the analysis because they were selected non-randomly. It also, sensibly, excluded 16 delegates who were assigned to the control group but mistakenly given web pages by a partner during implementation.

representatives better, and for the VNA as a whole, as the profile of delegates was raised in the public view.⁹

In this paper I have argued that the results presented by MST do not support their main conclusion that transparency has an adverse effect. The first reason for the difference in conclusion takes MST's findings at face value and focuses only on the interpretation of results for their regression equations using interaction terms. Interpreted properly, the results predict an increase in the number of *questions* and the percent of *critical* questions for more than 80% of delegates, and an aggregate positive effect of transparency. While these predicted positive effects of transparency are not statistically significant, neither are most of the predicted negative effects of transparency. Only 3 or 4 of 64 provinces have levels of *internet* penetration for which the models predict a significant negative impact of the *treatment*. Even this finding, however, appears to be an artifact of the methodology and group characteristics.

The second reason for challenging the “adverse effects of sunshine” is that the observed difference in behavior by delegates reflects differences before the experiment, rather than after—the observed patterns are consistent with mean reversion. More generally, I argued that *internet* penetration is itself a measure of transparency, and this interpretation is more sensible than as a measure of the “intensity of treatment.” The patterns observed are consistent with diminishing returns—the places where the predicted impact of transparency was positive are the very places that were most information-starved to begin with. In the VNA session after the experiment, the model predicts that the most active delegates, in both the *control* and the *treatment* groups, were in places with higher *internet* penetration, and this effect was higher for the group of delegates *treated* with the enhanced spotlight. If *internet* penetration is a measure of access to information, therefore, then not only is there no support for an “adverse effect of sunshine”, on the contrary MST's results suggest that sunshine works.

Finally, I showed that the key results on the post-treatment impact of transparency on the delegates are not robust. Of the three key findings on reelection and promotion prospects, one was the result of a programming error and the other two are sensitive to even a single influential observation. Taken as a whole, there is little reason to believe that anything negative befell the delegates who spoke out in the sunshine.

On a more constructive level, two methodological lessons can be drawn. First, interpretation of models with interaction effects should take due care to reflect the actual distributions of independent variables. In this case, taking account of the highly skewed and discrete distribution of the variable *internet*, and noting that it covers virtually the entire population, leads to a different set of conclusions than if these three facts are ignored.

A second lesson surrounds the drawing of samples for randomized experiments. It is self-evident that randomized experiments should draw *treatment* and *control* groups that are

⁹ Indeed, one may conjecture that the experiment helped boost the overall level of activity of the VNA, among both *treated* and *control* delegates. MST point out that although it is “thought to be commonplace that political activity is restrained in the year before the Congress, ... the sixth session of the VNA actually appeared [qualitatively] slightly more active than previous sessions,” and quantitatively “the session does not appear to be very different from previous sessions.” In this paper I have, for convenience, followed MST in the use of the term “treated” to refer to those on whom the spotlight was placed for their experiment. Just how cleanly one may separate the “treated” from the “control” delegates, however, is debatable. If the experiment influenced the behavior of both the *treated* and the *control* delegates, this would help explain why the VNA in the 6th session defied the prediction of reduced activity prior to the Congress. This is speculative, of course, and untestable. It does help illustrate, however, a key challenge of undertaking a randomized experiment in this way, as there is no placebo available to help isolate the effects of the treatment.

similar, and in fact MST attempted to do just this. If the proposed functional form for analysis, however, involves an interaction effect, then the pre-treatment relationships between the interacted variables should similarly be taken into account. Without doing so, there is the risk that changes in behavior could spuriously be attributed to the *treatment*, rather than to pre-treatment differences between the *treatment* and *control* groups. This is precisely what we observe with respect to the pre-treatment differences in MST's experiment.

MST's paper provides an interesting narrative of theories of decision making in an authoritarian country. Although the narrative at times paints a caricature that would put Vietnam at the extreme authoritarian end of some continuum of regimes, a characterization that does not sit well with the many other facts they cite¹⁰, MST's stories of control and retribution are nevertheless quite plausible. It is important, however, to distinguish between problems of freedom of expression and openness to debate on the one hand, and the impact of transparency on those problems on the other. The key issue tested in MST's paper is not whether dissent and criticism sometimes lead to negative consequences. Few would argue that this is not the case. Rather, the key issue, as highlighted in the paper's title, is whether enhanced transparency has adverse effects. I have argued that the evidence of such negative effects ranges from weak to non-existent, and that many of the results suggest a positive effect of sunshine.

Throughout this paper I have worked within the bounds of MST's theory and empirical framework, focusing only on empirical findings and interpretations. In closing, however, a more fundamental issue needs to be raised, and that is MST's assumption that the benefit of transparency is that it will make legislators more questioning and critical. Notwithstanding the impact MST's experiment may have had on delegate behavior, many would argue that the real benefits of transparency lie elsewhere, in the participation of the citizenry in public life, in the expansion of a fundamental freedom that is part and parcel of what it means to be developed (Sen 1999). The 1.3 million page views to the web sites created for MST's experiment attest to the fact that Vietnamese citizens, at least, place value on transparency.

¹⁰ These include the willingness of delegates to participate in the experiment, the existence of contested elections, the high-profile and televised nature of VNA debate sessions, the willingness of state media to carry out the experiment, etc.

References

- Abadie, Alberto (2005). "Semiparametric Difference-in-Differences Estimators." *Review of Economic Studies* 72: 1–19.
- Brambor, Thomas, William Roberts Clark, and Matt Golder (2006). "Understanding Interaction Models: Improving Empirical Analysis." *Political Analysis*. 14:63-82.
- Braumoeller, Bear F. (2004). "Hypothesis Testing and Multiplicative Interaction Terms." *International Organization*, Vol. 58, No. 4 (Autumn, 2004), pp. 807-820.
- Duflo, Esther, Rachel Glennerster, and Michael Kremer (2006). "Using Randomization in Development Economics Research: A Toolkit."
- Malesky, Edmund, Paul Schuler, and Anh Tran (2012). "The Adverse Effects of Sunshine: A Field Experiment on Legislative Transparency in an Authoritarian Assembly." *American Political Science Review*. Vol. 106, No. 4. November 2012. 762-786.
- Sen, Amartya (1999). *Development as Freedom*. Oxford University Press.
- World Bank and others (2009). *Vietnam Development Report 2010 – Modern Institutions*. Joint Donor Report to the Consultative Group Meeting, December 2-3, 2009. Hanoi.