

12 The impact of electoral debate on public opinions: an experimental investigation of the 2005 New York City mayoral election

*Sendhil Mullainathan, Ebonya Washington,
and Julia R. Azari*

I. Introduction

Political debates have long been a part of the American polity. The early practice was aimed primarily at promoting parliamentary and elite deliberation, as well as informing the broader public about issues and candidates (Jamieson and Birdsell 1988: 37).¹ Televised debates have altered this dynamic, structuring debate as a practice to educate the electorate. Yet the impact of televised debates on the quality of democracy has been somewhat controversial, with proponents casting debates as opportunities “to provide sustained analysis of issues and close comparison of candidates” (ibid.: 5), and others expressing the concern that the televised debates would increase the emphasis on image rather than substance (Druckman 2003).

Whether debates influence opinions by showcasing candidate viewpoints or by simply presenting shallow cues of candidate image, conventional wisdom assumes that televised debates will somehow influence viewers’ opinions regarding the candidates. Information in the electoral process is presumed to help voters make decisions more in line with their preferences (Lupia and McCubbins 1998). In light of the emphasis on the informative potential of debates, it is natural to pose the question of whether debates truly provide citizens with information that influences their opinions or choices.

To this end, there is a large literature examining the impact of debates on citizens’ political opinions and voting behavior. This research has concluded that debates are able to alter viewers’ opinions of candidates,

We are grateful to Tiffany Davenport and Meredith Levine for research assistance, Gregory Huber for his discussion, and ISPS for funding.

¹ The Lincoln–Douglas debates of 1858 were notable in this regard. They were well attended and addressed a number of issues related to the national question of slavery (Jamieson and Birdsell 1988: 49).

but only in a small way. This literature states that debates have more influence the earlier in the electoral season they occur. Further, the literature concludes that debates have greater influence over the uninformed than over those who know quite a bit about the election.²

Debate studies typically follow a panel design, interviewing voting-aged individuals both before and after the debate of interest. The pre-post-debate opinion change of those who watched is compared with the opinion change of those who did not watch in a “difference-in-difference” style methodology. The limitation of such an analysis is the potential endogeneity of debate watching. Perhaps only those who are most inclined toward opinion change actually watch the debate. In such a case the coefficient on “watch” would be an overestimate of the impact of the debate on the average citizen. Or perhaps only political junkies, who have the longest-held, most intransigent, political opinions, are in the viewing audience. In such a case the coefficient on “watch” would be an underestimate of the impact of the debate on the average citizen.

We overcome this limitation by employing an experimental panel design. During the 2005 election season, in the days leading up to the final debate between mayoral incumbent Republican Michael Bloomberg and Democratic challenger Fernando Ferrer, we interviewed a random sample of 1,000 New York City voters. We randomly assigned these 1,000 individuals to one of two groups: we asked the treatment group to watch the November 1 debate, and we asked the control group to watch a “placebo” program, PBS’s *The NewsHour with Jim Lehrer*, which aired opposite WNBC’s debate broadcast. Our intervention was effective. Statistically indistinguishable from controls in the pre-debate interviews, in post-debate interviews treatment group members reported watching the debate at a rate that was more than twice as high as controls. (We verified viewership by asking factual questions about the debate.) Thus we can take opinion differences between treatment and controls as a measure of the causal impact of the debate on political viewpoint.

Using this methodology we find that those in the watch group were 6 percentage points more likely to report that their opinions of one or both candidates had changed from the first to the second interview. However, when asked to rate the candidates on two of the most salient campaign issues, terrorism and housing, the post-debate responses of the treatment group were statistically indistinguishable from those of the control group. Clearly in a short survey we could not address every possible issue on which opinions may have changed. However, the fact that there was also no statistical difference between treatment and control group members’

² See Hellweg, Pfau, and Brydon (1992), Holbrook (1996), or Kraus (2000), for a review of this literature.

ratings of candidates generally, nor on their likelihood of voting for each candidate, strongly suggests that although the debate led respondents to *believe that their opinions had changed*, the debate actually effected *no meaningful political opinion change*. The study highlights the incongruence between respondents' stated opinion change and respondents' actual opinion change, a discrepancy pointed out by Gerber and Green (1999).

Why do we see no real opinion change in this study? Several possibilities exist: the first is that debates actually change no one's political opinions. Second, it is possible that debates shift everyone's political opinions. Because the highlights of debates are reported through the media and over the water cooler, perhaps one does not need to watch the debate to be affected by it.³ Third, perhaps debates on average do affect opinion change, and this debate was an exception. In other words, perhaps the external validity of the present work is limited. This could be due to its timing – just days before the general election – or perhaps because of the fact that this particular contest was never close. Or perhaps our sample – older, more educated and more likely to vote than the average New Yorker – is not a population whose views are swayed by debates.

In the next section we detail the history of the 2005 New York City mayoral election. In Section III we describe the data and methodology, before presenting results in Section IV. Section V discusses some potential implications of the findings in the context of the broader question of political knowledge and democracy. The final section concludes.

II. 2005 New York City mayoral election⁴

Former Bronx borough president, Democrat Fernando Ferrer, was never a serious contender in the 2005 New York City general mayoral election. The top vote-getter in the Democratic primary, he garnered only 39.949 percent of the vote. By law, 40 percent must be won by a single candidate to avoid a runoff. Fortunately for Ferrer, the second place finisher decided, for the sake of the party, to refrain from participating in the runoff. And thus Ferrer began his eight-week campaign against Michael Bloomberg, a Republican mayor, who enjoyed a 60 percent approval rating in an overwhelmingly Democratic city. (Seventy-four percent of the city's votes were cast for Democrat John Kerry in the 2004 presidential election [Andersen 2004].)

³ We turn to post-debate interviews conducted the night of the debate to control for this diffusion effect. Unfortunately, the sample interviewed that night was too small for this analysis to be informative.

⁴ The sources for this section are various articles from the 2005 *New York Times*: Healy (2005a–d); Anon. (2005); Purnick (2005); Healy and Lueck (2005); Healy and Connelly (2005); Haberman (2005); and Rutenberg (2005).

Even more than approval, Bloomberg had money. In what the *New York Times* regarded as “drowning-by-spending,” the billionaire incumbent spent more than \$70 million on his campaign. Ferrer’s only possible opportunity to present himself on an equal footing with the mayor came through the three debates endorsed by the city’s campaign finance program. But as Bloomberg was financing his own campaign, he was under no obligation to participate in these events. He decided to meet Ferrer in two of the three. Bloomberg avoided the first debate because he did not like its format: the October 9 event included Conservative Party candidate Thomas Ognibene who attacked Bloomberg for failing to be a true Republican. Ferrer and Ognibene spent most of the debate throwing criticisms at the empty lectern where the mayor would have stood.

The remaining debates occurred on October 30 and November 1, much later in the electoral season. At this point Bloomberg enjoyed a 27 percentage point lead in the polls. Bloomberg even led Ferrer amongst the more often Democratic subgroups of liberals, blacks and women. Nonetheless, at 9 a.m. on Sunday, October 30, Ferrer “tore into Mayor Michael Bloomberg.” Ferrer had what the *New York Times* dubbed “the best day so far in his race for mayor.” Two days later, however, Bloomberg evened the debate score. He was “far more assertive” than in the first debate in which he participated. The media and our survey respondents agreed. Bloomberg won this final debate, which is the focus of the inquiry in this chapter. The mayor then went on to win the November 9 election by more than 20 points, a record win for a Republican mayoral candidate in New York City.

III. Data and methodology

We investigate the question of how this final debate between Bloomberg and Ferrer affected public opinion of these candidates by employing an experimental panel design. Respondents, randomly selected from the New York City voter registration list,⁵ were first surveyed by telephone between October 28 and October 31. Using this initial survey we established respondents’ pre-debate attitudes about how life was going in New York City and about the two mayoral candidates. Respondents were randomly divided into one of two groups: we asked the treatment group to watch the final mayoral debate on November 1, and we asked the control group to watch a placebo program, PBS’s *The NewsHour with Jim Lehrer*, which aired opposite WNBC’s debate broadcast. We then interviewed

⁵ Our selection procedure gave greater weight to likely voters. The survey company phoned the home of likely voters but then asked for the adult registered in New York City whose birthday would come the soonest.

respondents for a second time between November 1 and November 3 to establish post-debate attitudes toward the candidates. Our final sample contains pre- and post-debate interviews with 1,000 New York City residents: 505 in the control group and 495 in the treatment condition.

Our methodology is simple. We measured the debate's impact on political views by examining the difference between treatment and control group members' post-debate views. Given our panel design, an alternative strategy would have been to compare the change (from pre- to post-debate) in political views of treatment and control group members. We did not follow such a strategy because we did not want to ask precisely the same questions in the pre- and post-debate interviews. Given that the pre- and post-debate interviews occurred within days of one another, we wanted to avoid respondents' post-debate answers being influenced (or anchored) by their pre-debate answers. Instead of this explicit differencing, we performed an implicit differencing strategy by assuming that members of the treatment group held, on average, the same political views as controls. Therefore, a comparison of post-debate opinion differences between the groups is equivalent to a comparison of pre-post-debate changes in opinion differences between the groups.

Thus the first assumption that underlies our identification strategy is that treatment and control group members do not differ significantly in political views before viewing the debate. We test this assumption in [Table 12.1](#) by comparing the two groups on pre-debate opinions.

The first column of the table gives the sample mean on the opinion (demographic). For example the 0.77 in the first column indicates that 77 percent of the sample agreed with the statement, "New York City is headed in the right direction." The second and third columns of the table give the means for the Watch (treatment) and Don't Watch (control) groups. These columns indicate that 78 and 75 percent of the Watch and Don't Watch groups respectively agreed that the city was headed in the right direction. In the final column of the table we provide the results of a test of the null hypothesis that the two groups are drawn from populations with the same mean. The fact that the final cell in the first row of the table is blank therefore says that the difference between the Watch and Don't Watch groups in opinion of whether the city is headed in the correct direction is statistically insignificant.

The same is true for the remaining pre-debate opinions. Control and treatment group members are statistically indistinguishable on party preference. The sample mean is 15 percent Republican and 75 percent Democrat. (By treatment status, those figures are 16 percent Republican and 73 percent Democrat for treatments and 13 percent Republican and 76 percent Democrat for controls.) The groups are also equally likely to choose Bloomberg (Ferrer) as their preferred candidate. (The figures are

Table 12.1 *Summary statistics*

	Full sample	Watch group	Don't Watch group	Test of equality of Watch and Don't Watch groups
<i>Pre-debate opinions</i>				
NYC is headed in the right direction	0.77	0.78	0.75	
Republican	0.15	0.16	0.13	
Democrat	0.75	0.73	0.76	
Will vote for Bloomberg (asked of only half the sample)	0.65	0.64	0.66	
Will vote for Ferrer (asked of only half the sample)	0.31	0.34	0.28	
<i>Voting history</i>				
Voted in 2001	0.96	0.96	0.97	
Voted in 1997	0.80	0.78	0.81	
<i>Demographics</i>				
Age	64	63	64	
White	0.71	0.68	0.74	**
Black	0.19	0.21	0.17	
Female	0.61	0.61	0.61	
College graduate	0.52	0.52	0.52	
<i>Post-debate checks</i>				
Watched debate	0.27	0.37	0.16	***
Correctly identified moderator's race (of 268 who claimed to have watched)	0.56	0.6	0.46	**
Correctly identified moderator's gender (of 268 who claimed to have watched)	0.64	0.69	0.54	**
N	1,000	505	495	

Note: *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level. Sample size varies due to item non-response.

64 percent for Bloomberg and 34 percent for Ferrer among treatments and 66 percent for Bloomberg and 28 percent for Ferrer among controls.) In summary, the evidence in the first section of [Table 12.1](#) supports the assumption that pre-debate opinions are the same for the two groups.

We further test this assumption by looking at voting history. By matching our sample to New York City voting records we examine whether respondents voted in the mayoral elections of 2001 and 1997. The 96 percent voting rate in 2001 reflects the fact that we predominantly sampled those who, based on their voting history, were likely to vote in the upcoming election. In neither year is the difference in voting rate between treatment and controls statistically meaningful. Thus voting history data provide further support for the assumption that pre-debate treatment and controls are the same politically.

Finally, we test the assumption by turning to demographic characteristics. On age, gender, and education there is no meaningful difference between the two groups. Only on one demographic, percentage of whites, do we see a statistically significant difference. The Watch group is 68 percent white; the Don't Watch group is 74 percent white. Clearly one statistically significant difference in 12 is no indictment of our randomization procedure. Nonetheless, we will present our results with and without controls for demographics to adjust for any pre-debate differences between the two groups.

Our assumption of equivalence of opinions among treatments and controls thus seems defensible. A second assumption on which our analysis rests is that our intervention – asking respondents to watch the debate⁶ – was in fact effective. In the final section of [Table 12.1](#) entitled “Post-debate checks” we test this assumption. The row marked “Watched debate” indicates that 37 percent of the treatment group but only 16 percent of the control group report having watched the debate. This difference is statistically significant at the 1 percent level. But our assumption is not that those in the treatment group report watching the debate at a greater rate than controls, which could arise out of a desire, on the part of respondents, to please the interviewer.⁷ Our assumption is that those in the treatment group actually watched the debate at a greater rate than controls. To test the veracity of respondents’ viewing reports, we asked those who claimed to have watched the debate to identify both the race and the gender of the moderator. The moderator of the November 1 debate was a white man, while the panelists (questioners) included a white man, a white woman, a black man and a Latino man.⁸ Fifty-six percent of those who claimed to have watched the debate identified the race of the moderator correctly. Sixty-four percent of those who claimed

⁶ Fifty-four percent of respondents asked in the pre-debate survey indicated that they would watch the debate for us. Thirty-one percent said they would not. The remaining respondents were not sure.

⁷ See, for example, Gales and Kendall (1957); Gray (1956); or Hanson and Marks (1958) on interviewer effects.

⁸ The moderator’s race and gender were the same for both debates in which Bloomberg participated. So unfortunately respondents may be given credit for correctly identifying the moderator when they are in fact correctly identifying the moderator for the wrong debate. Over 80 percent of the sample had their pre-debate interview before October 31 – the date of the first Bloomberg/Ferrer debate. Thus it is possible that some treatment members mistakenly watched the wrong debate. If half of those watching saw the debate that Bloomberg won and half saw the debate in which Ferrer prevailed, then it is possible that the reason we find no debate effects on average is because the debate in fact pulled the two groups in opposite directions. However, such a split in debate watching cannot explain the fact that on average debate watchers claimed to feel relatively more favorable to Ferrer rather than unchanged. In the debate in which the mayor did not participate the moderator was a black man.

Table 12.2 *Impacts of commitment to vote on voting (dependent variable is an indicator for having voted in the 2005 election)*

	(1)	(2)	(3)
Commit group indicator	0.042* (0.023)	0.043* (0.024)	0.040* (0.024)
Demographic controls		Yes	Yes
Voting history controls			Yes
N	795	721	721

Notes: Demographic controls include age, race, gender, and education. Voting history controls include indicators for voting in the 1997 and 2001 New York City mayoral elections.

to have watched the debate identified the gender of the moderator correctly. Thus it seems that respondents, on average, are truthfully reporting their viewership.⁹ Amongst those who claim to have watched the debate, treatment group members are significantly more likely to correctly identify the moderator's race and gender. To the extent that incorrect answers on moderator gender and race proxy for not actually having watched the debate, our results show that those in the treatment group were more likely to have watched.

The idea that simply phoning an individual and asking her to undertake some activity (in this case watching a debate) at a future date for no remuneration would result in compliance is hard to believe. Therefore we tested our methodology with an activity that we could actually verify – voting. In the post-debate survey, respondents, regardless of treatment/control debate status, were randomly assigned to a commit condition or a control condition. Those in the commit condition were asked an additional two questions at the end of the survey. The first was whether they would vote in the upcoming mayoral election (96 percent said they would) and the second was whom they would vote for (58 percent said Bloomberg). Those in the control group were not asked about future expectations for voting nor for vote choice. We then matched our sample to voting records. As the results in Table 12.2 indicate, those in the treatment group voted in the 2005 election at a rate that is a significant 4 percentage points higher than those in the control.

⁹ There is, however, the possibility that respondents learned of the moderator's demographics by seeing excerpts of the debate on the news. Fortunately the news outlets were more interested in giving sound bites of candidates responding than of the moderator's questioning.

Table 12.3 Summary statistics for commitment to vote treatment

	Commit group	Non-committed group	Test of equality of committed and non-committed groups
<i>Demographics</i>			
Age	65	62	*
White	0.71	0.72	
Black	0.21	0.18	
Female	0.63	0.59	
College graduate	0.50	0.54	
<i>Voting History</i>			
Voted in 2001	0.96	0.97	
Voted in 1997	0.82	0.77	
<i>Pre-debate opinions</i>			
NYC is headed in the right direction	0.77	0.76	
Republican	0.15	0.15	
Democrat	0.75	0.74	
Sample size	500	500	

Note: *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level. Sample size varies due to item non-response.

This is a sizable increase on a mean of 88 percent voting.¹⁰ As the final columns of the table show, this 4 percentage point difference is robust to the inclusion of controls for demographics and for voting history. The means for these demographics, by voting treatment status, are presented in Table 12.3. Thus this simple voting manipulation experiment provides evidence of the plausibility of the effectiveness of our debate manipulation.

IV. Results

The basic results of our survey are derived from the post-debate interview. As the results in Table 12.4 indicate, those in the Watch group were more likely than those in the Don't Watch group to believe that their opinion of one or both candidates had changed since the time of their initial interview.

¹⁰ Four percentage points is about half the expected size of the most effective GOTV (“get out the vote”) interventions, such as personal canvassing (Gerber and Green 2000).

Table 12.4 *Post-debate summary statistics*

	Full sample	Watch group	Don't Watch group	Test of equality of Watch and Don't Watch groups
Have experienced change in opinion toward one or both candidates	0.48	0.51	0.45	*
Opinion of Bloomberg changed	0.30	0.33	0.27	**
More favorable to Bloomberg	0.10	0.09	0.10	
Opinion of Ferrer changed	0.35	0.37	0.32	
More favorable to Ferrer	0.00	0.04	-0.03	
Bloomberg seems more knowledgeable	0.83	0.82	0.85	
Bloomberg thermometer	69	70	69	
Ferrer thermometer	48	49	47	
Bloomberg-Ferrer thermometer	21	21	21	
Bloomberg better on terrorism	0.89	0.89	0.90	
Bloomberg better on housing	0.53	0.50	0.55	
Will vote for Bloomberg (asked of only half the sample)	0.69	0.70	0.67	
Will vote for Ferrer (asked of only half the sample)	0.28	0.27	0.29	
<i>N</i>	1,000	505	495	

Note: *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level. Sample size varies due to item non-response.

The first row of Table 12.4 shows the mean answers to precisely that question. Forty-eight percent of the full sample said that their opinion had changed. By treatment status, 51 percent of those in the Watch group and 45 percent of those in the control group agreed with that statement. This 6 percentage point difference, statistically significant at the 10 percent level, is the intention-to-treat effect. However, we know that not everyone in the treatment group actually watched the debate. The difference in viewership between treatments and controls was not 100 percentage points, but merely 21 percentage points. Thus we scale our 0.06 effect size by 0.21 and conclude that the effect of the debate on those who actually watched was a 29 percentage point increase in the belief that their opinion had changed over the past few days. Or if we only consider those who correctly identified the moderator's demographics as having seen the debate, our treatment effect rises to a 43 percentage point increase in the belief that their opinion had changed.¹¹

¹¹ Fifty-seven percent of the 37 percent of treatments who said they watched the debate correctly identified the moderator's race and gender. Or in other words 21 percent (0.37×0.57) of treatments correctly identified the moderator's race and gender.

We next asked participants about opinion change regarding each of the candidates individually. Thirty-three percent of the Watch group felt their opinion of Bloomberg had changed; while only 27 percent of controls agreed with that statement, a difference that is statistically significant at the 5 percent level. This 6 percentage point intention to treat effect corresponds with a 29 percentage point treatment effect (or 43 percentage points if we consider only the verified viewership).

While those in the Watch group were more likely to say their opinion of Bloomberg had changed, they were no more likely than those in the Don't Watch group to say that their opinion of Bloomberg had changed in a positive (or negative) direction. The third row of the table examines responses to the question of how opinions toward Bloomberg had changed over the last few days. Respondents were offered the answer choices of "more favorable" coded 1, "unchanged" coded 0 or "less favorable" coded -1 . On average survey respondents felt more favorable towards Bloomberg, but there were no significant differences between the two groups. Watch group members saw a mean of 0.09 for this variable and Don't Watch members had a mean of 0.10.¹²

Watch group members were also more likely than controls to say that their opinion of Ferrer had changed. Watch group members expressed this view 37 percent of the time compared to only 32 percent of the time amongst controls. This 5 percentage point difference, however, is not statistically significant. Treatment group members on average saw their opinion of Ferrer rise, while controls saw their opinion fall. This 0.07 point difference (0.04 v. -0.03) is just shy of significance at conventional levels.¹³ Thus when asked directly whether their opinions had changed, those who were exogenously induced to watch the debate were more likely than controls to indicate that their opinions had changed in the last few days.

Yet when their opinions were measured directly, those in the treatment group held post-debate opinions that were statistically indistinguishable from those of treatment group members. The remainder of [Table 12.4](#) shows these results. The first direct opinion question (shown in row 6

Viewership of only 7 percent of controls (16 percent claimed to have viewed \times 43 percent of those named the moderator race and gender) was verified. Thus the difference between treatment and control verified viewership is 14 percentage points. We divide 0.06 by 0.14 to arrive at a treatment effect of 43 percentage points.

¹² This lack of significance between the two groups is caused not merely by an equal number of positive and negative Bloomberg switchers in the treatment group. The two groups saw comparable positive and negative change: 12 (20) percent of treatments had a less (more) favorable opinion of the mayor and 8 (19) percent of controls had a less (more) favorable opinion of the mayor.

¹³ The p value on the t test is 0.1067.

of the table) asked respondents whether they felt Bloomberg or Ferrer was more knowledgeable. Eighty-two percent of the Watch group and 85 percent of the Don't Watch group chose Bloomberg. The three-point difference is statistically indistinguishable. Next, we asked respondents to rate their feelings about each of the two candidates on a scale from 0 to 100.¹⁴ Once again, there were no significant differences between treatment and control group members. The groups rated Bloomberg 70 and 68 respectively and Ferrer 49 and 47. The average difference in rating between the two candidates, 21 points for both groups, was also statistically insignificant. We turned to two hot button issues of the electoral campaign next. We asked which candidate would perform better on (1) terrorism and (2) housing. Eighty-nine percent of Watch group members and 90 percent of Don't Watch group members thought Bloomberg would outperform Ferrer on the terrorism issue. On the housing issue, Bloomberg was favored by 50 percent of treatments and 55 percent of controls. Neither difference was statistically significant. Finally, we asked respondents for whom they planned to vote. Bloomberg was chosen by 70 percent of Watch group members and 67 percent of Don't Watch group members. Ferrer was chosen by 28 and 29 percent of the two groups respectively. Once again, neither difference was statistically significant. Thus the evidence of the second half of Table 12.4 indicates that the debates had no impact on political opinions.

The basic finding of an impact of debate watching on belief of opinion change but of no impact of watching on actual opinion change is robust to the inclusion of demographic and voting history controls, as shown in Table 12.5.

The first column of the table reproduces the results of Table 12.4 in regression form. Thus whereas in Table 12.4 we run *t*-tests, in Table 12.5 we test the equality of effect between the two groups by running regressions of the following form:

$$(1) \text{ Outcome} = a + B (\text{Watch group indicator}) + e.$$

We run a separate equation for each outcome of interest. As the tests are equivalent the results are of course identical: we see a significant impact of debate watching on the belief that someone's opinion has changed in general and that their opinion of Bloomberg has changed in particular. In column 2 of the table we add demographic controls (summarized in Table 12.1) to the model. Our qualitative findings remain unchanged. We continue to see significant impacts on changes of opinion of Bloomberg. Additionally, the finding that the Watch group has raised

¹⁴ The ordering of the candidates in the questioning varied randomly.

Table 12.5 *Impact of debate on opinions of candidates (each cell presents the coefficient on a different regression).*

	(1)	(2)	(3)
<i>Outcome</i>			
Have experienced change in opinion toward one or both candidates	0.058* (0.033)	0.057* (0.034)	0.049 (0.039)
Opinion of Bloomberg changed	0.06** (0.029)	0.053* (0.031)	0.055 (0.035)
More favorable to Bloomberg	-0.011 (0.035)	-0.023 (0.037)	-0.014 (0.041)
Opinion of Ferrer changed	0.051 (0.031)	0.053 (0.033)	0.044 (0.037)
More favorable to Ferrer	0.062 (0.039)	0.082** (0.041)	0.075* (0.046)
Bloomberg seems more knowledgeable	-0.032 (0.027)	-0.033 (0.028)	-0.033 (0.031)
Bloomberg thermometer	1.234 (1.731)	0.870 (1.798)	-0.774 (1.925)
Ferrer thermometer	2.117 (1.721)	20.081 (1.790)	0.987 (1.974)
Bloomberg–Ferrer thermometer	-0.758 (2.412)	-1.344 (2.488)	-2.341 (2.728)
Bloomberg better on terrorism	-0.001 (0.022)	0.007 (0.024)	0.016 (0.025)
Bloomberg better on housing	-0.045 (0.035)	-0.033 (0.036)	0.000 (0.040)
Will vote for Bloomberg (asked of only half the sample)	0.028 (0.045)	0.025 (0.046)	0.028 (0.050)
Will vote for Ferrer (asked of only half the sample)	-0.013 (0.044)	-0.002 (0.045)	-0.009 (0.049)
<i>Controls</i>			
Demographic controls		Yes	Yes
Voting history controls			Yes

Note: Sample size varies among specifications due to non-response. Total sample size is 1,000. *** denotes significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

its opinion of Ferrer more than the control group has is now statistically significant at the 5 percent level. All of the direct measures of opinion change continue to show insignificant coefficients on the treatment status dummy. In column 3 we add voting history controls (also summarized in Table 12.1) to the model of column 2. Coefficients are little changed for all outcomes as we move from column 2 to column 3. However, as we utilize increasing degrees of freedom, the general opinion change and

the Bloomberg opinion change specifications no longer show significant treatment/control differences. The only significant difference between the groups remains in their opinions of Ferrer. The treatments believe, more than controls, that their opinion of Ferrer has improved over the last few days.

Thus our basic finding is robust to the inclusion of demographic and voting history controls: we show that debate watching increased participants' beliefs that their opinions of the candidates had changed in the past few days. However, when measuring directly, we find no evidence of actual opinion change.

V. Implications

Since the experimental design used in this study overcomes the endogeneity issues endemic to observational studies of debate effects, we cannot attribute the lack of change in participants' opinions to their preexisting strong opinions (Zaller 1992), or to some other shared characteristic of debate watchers. Therefore, the findings suggest that, independent of the effects of existing views or political awareness, the experience of viewing political debates did not change opinions about the candidates. In this section we discuss the implications of both this finding and the finding that respondents thought their opinions had changed when in fact they had not.

Democratic ideals depend on an informed electorate capable of making reasoned political choices (Delli Carpini and Keeter 1996: 1). Responsible party government enforced through electoral accountability requires mechanisms by which citizens can learn about the policy positions of the candidates vying to be their leaders. More candidate-centered approaches to political competition depend on such processes as well in order for citizens to make judgments about the character or personal qualities of candidates. As stated in the introduction, debates in the contemporary era have been seen as a means of helping candidates communicate their message to citizens and, in turn, for citizens to use the information gained by watching debates to form opinions of candidates (Jamieson and Birdsell 1988). In light of the intended relationship between debates and popular rule, the results of this experiment are particularly striking: we find that debates do not perceptibly change citizens' opinions, and – even more notable – despite the evidence that actual views had not changed, participants in this study were under the impression that their views had changed after watching the debates.

American political thought and history have long grappled with the tension between the ideal of popular rule and the possibility of a poorly

informed electorate. While revolutionary ideals may have rested in part on an image of “the people” who are “wiser than their governors” (Morone 1998: 5), a streak of skepticism about this characterization of “the people” runs through early American political thought as well. Madison repeatedly warns in the *Federalist* against the potential for the people to rule using “passion” rather than “reason.”¹⁵ While the 1776 Revolution against British rule was based on a radical democratic ideology, the dominant sentiment among ruling elites in the new American republic soon took a more conservative turn, motivated in part by a heavy skepticism about the ability of the mass electorate to make informed political decisions (Wood 1998).

The question of whether popular rule is advisable or even possible under conditions of limited citizen information is very much alive in contemporary normative and empirical studies. Contemporary inquiries into the role of information in the functioning of American democracy call into question the level of information possessed by the American electorate (Delli Carpini and Keeter 1996), the effect of political awareness and reception of political messages on public opinion (Zaller 1992), and the importance of information for making reasoned political decisions (Lupia and McCubbins 1998).

How does the literature on political knowledge bear upon the implications of a study whose findings concern the *opinions*, rather than the political knowledge, of the participants? Implicit in the literature on political knowledge is a link between information and political judgment and behavior. Lupia and McCubbins, who investigate this link systematically, insist that perfect information is not needed for citizens to make reasoned political choices; what is required for reasoned choices is what the authors call “knowledge” (ibid.: 6), defined as “the ability to predict the consequences of actions.” Their point speaks directly to the implications of this study: the purpose of practices intended to enhance political knowledge is to give citizens the tools to change their opinions about candidates in order to more closely match their true preferences. Our discussion of the implications of the results of the study in this broader context must be cautious; we do not know anything for sure about the voting behavior of the study participants. Obviously we cannot say anything about the ultimate impact of the perceived shift in opinions following the debates.

¹⁵ Madison *et al.* (1982: 310–14). The arguments advanced in these documents center on the concern about creating institutional arrangements that will be conducive to the “people’s rule” using reason rather than passion, distinct from more contemporary ideas about creating an informed populace.

We can, however, cautiously consider the implications of the findings for the complex constellation of relationships at stake in the study of mass opinion and knowledge, media events such as televised debates, and broader notions of democratic rule. The fundamental purpose of debates – namely, to inform citizens about candidates’ characteristics and issue positions – has been explicitly linked with the objective of changing opinions; contemporary models of political information do not regard such information as an intrinsic good, but as means to an end. That end is the ability of citizens to make political choices that better reflect their underlying values and preferences. Here we return to Lupia and McCubbins’s thesis that reasoned political decisions are possible with limited political information as long as meaningful cues are available, and that political information is costly and “provides no instrumental benefit” if it does not change citizens’ decisions (*ibid.*). As stated before, we do not know whether watching the debate had an impact on citizens’ vote choice.

However, the results suggest that if opinions of the candidates are the bases of such decisions, we might expect that the debates did not have this kind of impact. Further research is needed to determine whether the stability in opinions after watching the debates is associated with stability in electoral behavior. The absence of substantial opinion shift following the debates may suggest that debates are costly and provide little actual value to citizens who rate political information only insofar as it helps them make political decisions that reflect their preferences and views. The fact that participants thought their opinions had changed when in fact they had not suggests that the process of incorporating new information may be more complex than previously theorized. For instance, we do not know whether the perceived changes in opinion of Bloomberg among the treatment group led to changes in voting behavior. However, the results of this study suggest that there may be a distinction between voting based on an actual opinion, and based on a perceived opinion. Our findings suggest that the relationship between political information, political opinions, and candidate evaluation may be even more complex than current scholarship indicates. Debates and other means of informing citizens about candidates may not function in as straightforward a manner as conventional wisdom would predict.

VI. Conclusion

Previous research on the impact of candidate debates on public opinion has suffered from potential bias due to the endogeneity of debate viewing. We overcome this limitation by employing an experimental panel design. We randomly assign study participants to watch either the final

2005 New York City mayoral debate or a placebo program. We find that debate watching increased participants' beliefs that their opinions of the candidates had changed in the past few days. However, when measured directly, we find no evidence of actual opinion change. Thus this study highlights the incongruence between respondents' stated opinion change and respondents' actual opinion change, a discrepancy pointed out by Gerber and Green (1999).

We do not conclude, however, that debates never induce opinion change. There are several possible explanations for our lack of opinion change findings in this context. The first is, of course, that debates actually change no one's political opinions. The second is that debates move everyone's political opinions. Because the highlights of debates are reported through the media and over the water cooler, perhaps one does not have to watch the debate to be affected by it. Third, perhaps our findings are not generalizable to the voting population at large. Our research setting – New York City – is one of the most politically homogeneous locales in the nation. Our focal debate occurred just days before the general election. Our sample was older, more educated and more likely to vote than the average New York City resident.

Thus we conclude with a suggestion and a word of caution. We suggest that future work on the persuasive effects of debates be conducted using an experimental research design. Such a design overcomes the bias from the endogeneity of debate viewership. We also caution against the use of stated opinion change as a proxy for actual opinion change. Our results highlight that these two variables do not measure the same underlying construct.

BIBLIOGRAPHY

- Andersen, Kurt. 2004. "People Like Us." *New York* 37 (November 22): 24–6.
- Anon. 2005. "The Shunned Debate." *New York Times*, October 9, p. CY11.
- Delli Carpini, Michael X. and Scott Keeter. 1996. *What Americans Know About Politics and Why it Matters*. New Haven, CT: Yale University Press.
- Druckman, James N. 2003. "The Power of Television Images: The First Kennedy–Nixon Debate Revisited." *Journal of Politics* 65 (May): 559–71.
- Gales, Kathleen and M. G. Kendall. 1957. "An Inquiry Concerning Interviewer Variability." *Journal of the Royal Statistical Society*, ser. A. (General) 120 (2): 121–47.
- Gerber, Alan, and Donald Green. 1999. "Misperceptions about Perceptual Bias." *Annual Review of Political Science* 2 (June): 189–210.
2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94 (September): 653–63.

- Gray, Percy G. 1956. "Examples of Interviewer Variability Taken from Two Sample Surveys." *Applied Statistics* 5 (June): 73–85.
- Haberman, Clyde. 2005. "Watch Debate? Sleep In? Boxers? Briefs?" *New York Times*, October 28, p. B1.
- Hamilton, Alexander, James Madison, and John Jay. 1982. *The Federalist Papers*. New York: Bantam Books.
- Hanson, Robert H. and Eli S. Marks. 1958. "Influence of the Interviewer on the Accuracy of Survey Results." *Journal of the American Statistical Association* 53 (September): 635–55.
- Healy, Patrick. 2005a. "Weiner Concedes Race for Mayor to Avert Runoff; Says Ferrer is Nominee." *New York Times*, September 15, p. A1.
- 2005b. "Ferrer Makes His Candidacy Look Stronger Than the Polls Suggest." *New York Times*, October 31, p. B4.
- 2005c. "Attacks Escalate in Final Debate of Mayor's Race; Rivals Go on Offensive." *New York Times*, November 2, p. A1.
- 2005d. "Bloomberg Cruises to Re-Election Victory; Corzine is Winner in Costly New Jersey Race." *New York Times*, November 9, p. A1.
- Healy, Patrick, and Marjorie Connelly. 2005. "Even Among Democratic Voters, Poll Finds Ferrer is Well Behind." *New York Times*, October 28, p. A1.
- Healy, Patrick, and Thomas Lueck. 2005. "Ferrer Levels Attack on Bloomberg Over Iraq War and the Need for Low-Cost Housing." *New York Times*, October 27, p. B3.
- Hellweg, Susan A., Michael Pfau, and Steven R. Brydon. 1992. *Televised Presidential Debates: Advocacy in Contemporary America*. New York: Praeger.
- Holbrook, Thomas M. 1996. *Do Campaigns Matter?* Thousand Oaks, CA: Sage Publications.
- Jamieson, Kathleen Hall, and David S. Birdsell. 1988. *Presidential Debates: The Challenge of Creating an Informed Electorate*. New York: Oxford University Press.
- Kraus, Sidney. 2000. *Televised Presidential Debates and Public Policy*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lupia, Arthur and Mathew D. McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Really Need to Know?* New York: Cambridge University Press.
- Madison, James, Alexander Hamilton, and John Jay. 1982. *The Federalist Papers*. New York: Bantam, 1982.
- Morone, James A. 1998. *The Democratic Wish: Popular Participation and the Limits of American Government*. New Haven, CT: Yale University Press.
- Purnick, Joyce. 2005. "So Far, a Race Without Chutzpah." *New York Times*, October 20, p. B1.
- Rutenberg, Jim. 2005. "In Equal Footing of a Debate, Ferrer Gets Feisty." *New York Times*, October 31, p. A1.
- Wood, Gordon S. 1998. *The Creation of the American Republic, 1776–1787*. Chapel Hill: University of North Carolina Press.
- Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.