

**Who Moderates the Moderators?
The Effect of Non-neutral Moderators in Deliberative Decision Making**

Paolo Spada (Harvard Kennedy School)
and
James Raymond Vreeland (Georgetown University)

October 2013

Abstract: Ideal deliberative democracy seeks to employ unbiased moderators. Yet, a large literature in the field of mediation suggests the elusiveness of perfect neutrality. Our study thus addresses the following question: when moderators of deliberations express their own views – even in a limited manner – can they change the preferences of participants? Using a novel experimental design in a real deliberative decision-making process, we find that moderators can significantly influence the attitudes and behaviors of participants by expressing views in a constrained manner. The results of our study have implications for refining epistemic conceptions of deliberative democracy and for designing more precise empirical investigations of the effects of deliberative processes on attitudes and behavior. The results also warn of a simple mechanism by which interest groups might hijack the deliberative decision-making processes used in community driven development projects all over the world.

Acknowledgements: For specific suggestions, the authors wish to extend their gratitude to Marc Busch, Daniel Butler, Jon Butler, Raj Desai, Franziska Fiebig, John Fulwider, Archon Fung, Alan Gerber, Donald Green, Zulia Gubaydullina, Gregory Huber, Macartan Humphreys, Kosuke Imai, Holger Kern, Stephan Klasen, Anna Maria Mayda, Quinton Mayne, Abraham Newman, Jennifer Tobin, and the anonymous reviewers. The authors also thank participants in the Mortara Center Tuesday Political Economy Lunch Group, the Graduate Student Conference on Experiments in Interactive Decision Making at Princeton University, and the Economics Department Seminar at Goettingen University. We would also like to acknowledge our moderators, whose generous assistance on this project was essential for our research: Onur Bakiner, Laia Balcells i Ventura, Sheree Bennett, Edwin Camp, Min Young Han, Dessislava Kirilova, Dominika Koter, Meghan Lynch, Melissa Mason, Shivaji Mukherjee, and Robert Person.

1. Introduction

Ideally, the procedures governing deliberative democracy – where public deliberation leads to a binding decision – should promote the inclusion of all participants in the discussion. These procedures should prevent disruptive behaviors and induce participants to listen to each other, allowing for as much diversity of opinion as possible. No single opinion should be privileged beyond the merits of its inherent rationality and logical appeal.¹

To achieve this set of normative goals, many practical applications of deliberative democracy divide the participants into small discussion groups of five to ten people, and assign moderators to facilitate the discussion. The ostensible goals of the moderators are simply to stimulate discussion and guarantee that all participants have the opportunity to speak – they are thus introduced as neutral interveners.

The procedures of deliberative democracy thus appropriately endow moderators with coercive power so that they can prevent domination by vociferous participants and thereby induce higher-quality group discussion.² Yet, the actual extent and form of moderator intervention varies significantly. Real-world applications of this approach include participatory budgeting and community-driven development projects used in emerging markets and developing economies throughout the world. In these contexts, moderators sometimes intervene in discussions by summarizing or reframing the opinions of participants, providing additional information, or even explicitly supporting one option over another.³ Our question is the following: What happens to the preferences of participants when moderators intervene in the discussion in support of one view over another?

Raising this question highlights an often overlooked fragility of real-life deliberative decision-making processes. Most critiques of empirical applications of deliberative democracy focus on the disruptive effect of inequality among the participants, the biasing effect of outside interest groups and media on deliberation, and the polarizing effect of group dynamics.⁴ Even in the absence of these problems, however, moderators themselves might bias the outcomes of

¹ In defining deliberative democracy, we follow Mansbridge *et al.* (2010), who posit that procedures should provide for both deliberation and a binding decision-making mechanism. For recent reviews of the literature see Bächtiger *et al.* (2010) and Elstun (2010). On the epistemic benefits of cognitive diversity in deliberative democracy, see Page (2007), and Landemore (2012). Our understanding of deliberative democracy has been particularly influenced by Ackerman (1991, 1998), Austen-Smith and Feddersen (2006), Barber (1984), Benhabib (1992, 1996), Besson and Martí (2006), Bohman (1996, 1998, 2009), Chambers (1996, 2003), Cohen (1986, 1989, 1996), Dryzek (2000, 2010), Elster (1997, 1998), Estlund (1997), Fishkin (1995, 2009), Fung and Wright (2003), Fung (2007), Gastil (1993, 2000), Gaventa (2006), Goodin (2005, 2008), Gutmann and Thompson (1996, 2004), Habermas (1984, 1987, 1996), Mansbridge (1980, 1999), Mansbridge *et al.* (2013), Nino (1996), Pateman (1970), Shapiro (2003), Warren (1992), Warren and Pearse (2008), and Young (1996, 1999, 2001).

² See Karpowitz and Mansbridge (2005), Levine, Fong, and Gastil (2005), Pierce, Neeley and Budziak (2008), and Mansbridge *et al.* (2010).

³ For a review of the variety of moderator techniques that have been developed in the field of mediation, see Boulle (2011).

⁴ See Sanders (1997), Shapiro (1999, 2002), Sunstein (2000), Young (2000), Mendelberg (2002), and Morrison and Singer (2007).

deliberative democracy. In real-world applications of deliberative democracy, moderators rarely hold neutral positions. In particular, if moderators have a stake in the outcome of the deliberative decision, they have an incentive to employ their coercive powers in ways that violate ideal deliberation. One of the authors of this study participated in the weekly meetings of the Participatory Budgeting Council in Porto Alegre, Brazil, observing moderators attempting to influence the outcome of discussion via interruptions, shouting, strategic agenda setting, and even direct threats towards participants. Now, academically-inspired experiments – such as Deliberative Polls®, citizens’ juries, and America Speak – employ neutral moderators, but thousands of participatory-budget deliberations and community-driven development projects employ moderators who can express explicit viewpoints.

This study thus tests the effect of moderators in a controlled field-experiment setting, where the views of moderators are randomly assigned to different discussion groups. We asked the participants (107 students of a political science class) to indicate their preferences over the number of writing and exam requirements for the course – issues that they cared about. They then deliberated in groups of about five students each. In some groups, the moderator expressed neutral views, while in others the moderator advocated for one option over another. After deliberating, the students cast final votes. The outcomes of the elections were binding on all students. We stress the binding nature of the decision because it is a rare feature in deliberative and participatory experiments. Many deliberative processes are only consultative, but in many high-stakes cases, the process is binding. We also stress that the participants were not aware that they were partaking in a study and thus behaved as they naturally would in a typical deliberative setting.

The experiment was implemented in the fall of 2007 with the approval of the Yale University Human Subjects Committee.⁵ The balanced nature of our experiment helped to ensure that the final results of the deliberative process would not be impacted – indeed, they were not. Moreover, the vast majority of the students reported that they favored the opportunity to participate in the decision-making process for the course requirements. Most of them recommended that we continue to use the process with future classes. No student expressed objection in the anonymous evaluations or in person.

As expected, the results of our experiment show that when moderators intervene in favor of one option, they can have a substantively and statistically significant impact on the preferences of participants. For each decision in our experiment (writing and exam requirements), participants could choose between two options (A or B) or declare indifference. Consider the basic results of the first discussion (the writing requirement).

In the control group (where moderators exerted no influence), 83 percent of the students voted for option A (3 short response papers), while 17 percent voted for option B (1 short response paper and 1 final paper); no student was indifferent. In groups where moderators influenced toward option A, the average treatment effect appears to increase the number of students in favor of option A by 14 percentage points. In groups where the moderators influenced

⁵ The study qualified for an exemption under 45 CFR 46.101(b)(2). See: <http://www.yale.edu/hsc/>. The exemption letter is available on request.

toward option B, the average treatment effect of treatment B appears to increase the number of students in favor of option B by 20.5 percentage points.

Similar results emerge from the second discussion (the exam requirement). In the control group 6.1 percent of the students voted in the referendum for option A (cumulative final exam), and 93.9 percent voted for option B (midterm examination and non-cumulative final examination); no student was indifferent. In groups where moderators influenced toward option A, the treatment appears to have reduced the number of students in favor of option B by 19.8 percent. Where moderators influenced toward option B, no student voted for option A, and the treatment apparently reduced the number of students in favor of option A by the full 6 percentage points.

Of course, in our setting, both of our discussions had unexpectedly large majorities in favor of one option over the other, which limited the effect that our moderators could have when influencing towards the popular option. So, the effectiveness of the moderator-effect can be more precisely estimated when considering cases where the moderator influenced towards the less popular option as we show below and in appendix 3.⁶

The results of our study have implications for refining epistemic conceptions of deliberative democracy and for designing more precise empirical investigations of the effects of deliberative processes on attitudes and behavior. The magnitude of our results implies that empirical studies investigating the effect of deliberation on attitude formation and behavior may need to disentangle the effect of moderators from the effect of deliberation. Without properly understanding the role of moderators, we might overestimate or underestimate the effect of deliberation. We return to this issue in the conclusion.

Additionally, this article details a relatively inexpensive methodology that can be applied to investigate other elements of the deliberative decision-making processes – from subtle moderator-interventions (including the use of body language), to the impact of providing information packages to participants, framing messages, and group composition. Future work could, for example, test for the separate effects of a moderator endorsement of one option over the other and a rational argument. For example, an excellent recent laboratory experiment by Park (2012) builds on our methodology to investigate the effect of moderators with different levels of expertise, finding, indeed, that greater expertise leads to more influence. Note that our treatment is composed of both a moderator endorsement combined with a moderator argument. Our experiment is not designed to separate these two components. Nor is the experiment designed to study the strength of different types of persuasive messages. A large literature in

⁶ Our experimental results measure the difference in the average post-discussion opinions between a treatment group, in which moderators expressed their views, and a “no influence” group, in which moderators did not (see section 3 for a detailed discussion and appendices 2 and 3 for a more detailed presentation of the data). By construction, the distribution of post-discussion votes in the control group dictates the upper bounds of the treatment effect and the possibility of distinguishing the effect of moderators from the effect of group discussion – that is, the precision of our estimate. In an extreme case, if there were a 100 percent majority in the control group, it would be impossible to detect any positive moderator’s effect influencing toward the majority option. Thus it is not surprising that when we study the effect of the moderators influencing toward the more popular option, we find smaller and less statistically significant results than when we study the effect of moderators influencing toward the less popular option.

social psychology investigates such issues in non-deliberative settings, and our methodology could be employed to test the conclusions of this literature as they apply to deliberative democracy.⁷

As the first study to randomly assign moderator-preferences within a deliberative decision-making process, our objective is to simply establish whether moderators can impact the preferences of participants with an intervention that is similar to – although more conservative than – those we have observed in real-world settings. The most interesting feature of our methodology is that the participants behave more naturally than in laboratory experiments.

As for practitioners, who employ the tools of deliberative democracy, our results flag the importance of moderators as a potential vulnerability that interest groups may seek to exploit. These phenomena are most relevant in community-driven development projects that allocate significant resources. If a reasonable degree of disagreement existed in a community, and organized interest groups had influence over moderators, they could hijack a deliberative decision-making process.

The rest of our article proceeds by first discussing the experimental literature on deliberative democracy (section 2). We introduce the details of our experimental methodology in section 3. We then present and discuss the results in sections 4 and 5, respectively. Section 6 concludes with a discussion of the implications of our research for the study and implementation of deliberative democracy.

2. Experimental studies

Scholars have conducted experiments to test the impacts of various facets of deliberative discussions on the post-discussion preferences of participants.⁸ Some researchers, for example, find that the effect of deliberation is smaller for salient issues and when the participants are better informed (Farrar *et al.* 2010, Luskin, Fishkin, and Hahn 2007). Others show that deliberation makes the preferences of participants more structured and orderly (List *et al.* 2006).

Some studies have tested the effect of inequality among the participants on the deliberative outcome. Morrison and Singer (2007) find that the perception of inequality among the participants influences whether participants find the outcomes of deliberation to hold legitimacy. Other studies have investigated the effect of group pressure and group composition on participants' behavior. Farrar *et al.* (2009), List *et al.* (2006), and Luskin *et al.* (2007) all find that participants modify their preferences to conform to the preference of their group. Their results contrast with the psychology literature that analyzes group polarization effects (see Sunstein 2000). Karpowitz, Mendelberg, and Shaker (2012) study the voice of women in a deliberation, finding a substantial gender gap biased against women (unless there are many women or a unanimous rule is employed in settings with few women).

⁷ For an overview of this literature see Cialdini (2007).

⁸ In our review, we focus on the experimental literature, but important non-experimental applications of participatory and deliberative democracy have also been pursued in business, law, medicine, game theory, and political science. For reviews, see Chambers (2003) and Mendelberg (2002).

Sophisticated research is also emerging on strategic behavior in deliberations. In a non-experimental study, Muhlberger (2007) finds that participants are minimally strategic. But several studies disagree with this conclusion (see, for example, Landa 2005; Hafer and Landa 2005, 2007; Meirowitz 2007; Dickson, Hafer and Landa 2008a,b; and Landa and Meirowitz 2009). For a recent study that combines game-theoretic and psychological approaches tested with an experimental design, see Myers (2010).

A number of studies of deliberation have focused on the role of moderators.⁹ Pierce, Neeley, and Budziak (2008) find that neutral moderators, who do not express their views during a deliberation, increase perceptions of fairness in the procedure, confirming the important role that moderators can ideally play. In his work, Steiner (2012) considers a rich set of experiments and case studies and concludes that the style of moderator-intervention impacts the ways participants contribute to discussions. In an experiment that randomly assigns groups to deliberate with and without a moderator, Fulwider (2005) finds that the presence of moderators makes opinion change more likely. As mentioned above, Park's (2012) laboratory experiments show that moderators perceived as having more expertise have more persuasive effects on participants.

Similarly, Humphreys, Masters, and Sandbu (2006) find that group responses are correlated with the preferences of moderators. They analyze the results of a national deliberation on country-wide economic priorities organized by the UN Development Program in São Tomé and Príncipe. The moderators were randomly selected from a group of civil society leaders and public officials, with some moderators randomly assigned to more than one group. The authors infer the importance of moderators from the share of the variance explained by moderator-specific effects. Moderator effects account for over one third of variation in the outcomes. The authors also provide evidence that opinions changed in the direction of moderator influence. Moderator preferences – which were not manipulated by the experimenters – were recorded one week after the deliberation and may thus have been influenced by the group discussions themselves, but this is unlikely as it would imply strange moderators who are both highly influential and highly malleable (Humphreys *et al.* 2006:598).¹⁰

Still, in contrast to these studies, which all detect various moderator effects, Farrar *et al.* (2009) find no robust differences in the policy views of participants across groups with moderators employing passive and active approaches (although in this experiment, moderators did not express their views).

None of the above studies randomly assign preferences to moderators in a field experiment setting. What happens when we manipulate the views of moderators in a natural setting? Can moderators influence participants to change their views when they argue for one option over another – even if they do so in a rather limited manner? Our experiment is designed to address these questions directly by scripting the intervention of moderators. It is the first study to

⁹ Among others, see Sanders (1997), Young (2001), Shapiro (2003), Karpowitz and Mansbridge (2005), Levine Fong, and Gastil (2005), Mutz (2008), Mansbridge *et al.* (2010), and Steiner (2012).

¹⁰ See Imai and Yamamoto (2010), however, for a critical discussion.

randomly assign the direction of moderator-influence and the only existing study to do so in a field experiment.

3. The deliberative decision-making process and the experimental design

The experiment took place in an Introduction to Comparative Politics course at Yale University with an enrollment of 107 students. The subjects were not aware they were participating in an experimental study.¹¹ The course syllabus noted the following: “After studying elections, we will be holding elections in class ourselves! Students will be able to vote for alternative writing requirements and alternative exam requirements. Majority rules.” The syllabus provided options for the writing and exam requirements (see table 1). Note that the outcomes of the elections were binding on all students in the class, thus the experiment represents a binding deliberative decision-making procedure.

Table 1: Writing and exam options

Writing requirement	Exam requirement
A) 3 short response papers (maximum 3 pages each)	A) Cumulative final exam (45%).
B) 1 short response paper (maximum 2 pages) and 1 final paper (paper should be 8 pages)	B) Midterm examination (covering lectures 1-15) and final examination (covering lectures 17, 19-25).

At the beginning of the semester, each student selected one of eight possible section-meeting times. During the fifth week of class, students were randomly assigned to groups within their respective section-meetings. There were a total of 24 groups with an average group size of 4.5 students.¹²

Before the group discussions took place, we conducted a confidential pre-survey to record the initial preferences of each participant. Students then discussed the writing requirement for 15 minutes in their groups before casting their secret ballot on the issue. Next, they discussed the exam requirement for 15 minutes before casting their secret ballot on this issue. A discussion rule was also randomly assigned to each group. Some groups had no rules, while others had “strict discussion rules,” which limited students to speaking for one minute at a time and required all students to have the opportunity to speak before a student could speak twice.

At the end of the 30-minute session, the students filled out a survey, which asked standard questions about the participants’ characteristics and to evaluate their perceptions of the process. In particular, we collected data on gender, graduation class, and membership in a varsity team, which we address in the analysis below.

¹¹ This is a unique feature of our field experiment – for which we sought out the specific approval of the Yale University Human Subjects Committee. In most field experiments, participants know that they are being studied.

¹² More precisely, six section-meetings had three groups, one had two groups, and one had four groups. There were nine groups of five students, one group of six students, and 14 groups of four students.

Note that because the writing and exam requirements were discussed and voted on separately, we observed a total of 48 separate group discussions (two for each of the 24 groups). In the results section below, we place more confidence in the results from the first discussion because the second discussion may have been contaminated by the first. In fact, however, the qualitative results hold across both discussions and are quantitatively similar.

Each of the groups had one moderator who was introduced as a graduate student.¹³ We randomly assigned moderators to intervene in discussions by expressing opinions (or not) over the various options. “Preferences” were randomly assigned as entire scripts per discussion. So moderator preference was consistent throughout a single discussion. A given moderator could, however, be neutral (placebo) in the first discussion and biased (treatment) in the second, and vice-versa. We scripted the interventions and report exactly what the moderators said for the first discussion in appendix 1 (the script for the second discussion is available on request). “Treatment” moderators made three statements supporting their “preferred” option. The first statement was read at the beginning of the discussion; the second, halfway through the discussion; and the third at the end. The fact that the moderator would intervene in the discussions was unknown to the students *ex ante*.

Other than reading the script, moderators assigned to influence the students could only attempt to influence the discussion using body language, nodding their heads during the discussion when students raised points that agreed with their “preference,” shaking their heads when students disagreed. Moderators assigned to be neutral also read three statements at exactly the same time intervals as the moderators assigned to influence the students, but these moderators expressed no opinion about the options and did not use body language.

The students were instructed to justify the preference they had written in the pre-survey during the discussions. This type of deliberative exercise, which stresses justification and critical evaluation, is used in many deliberative discussions. Additionally, we reminded the students that the decision was important because it would affect their coursework – the results of the referenda were binding on the entire class.

Regarding our randomization procedure, we did not generate significant imbalances in treatment assignment (see appendix 2). The pre-discussion preferences of the participants do not significantly predict treatment assignment, nor do any other individual characteristics that we investigated with the exception of membership in a varsity team. We flag this factor because varsity practice-times influenced the schedules of athletes, as a group, which influenced their section-meeting choice and may have also influenced their preferences on course requirements. We therefore present robustness checks that control for this factor in our analysis of the data below.¹⁴

¹³ We employed a total of 12 moderators. They each moderated between two and five of the 24 total groups. We tested for moderator-specific effects and found none, likely due to the scripted nature of the moderator-interventions.

¹⁴ Mutz and Pemantle (2011) argue against using randomization checks as a justification to introduce pre-treatment covariates in the regression analysis of experimental results. They recommend using simple regression and non-parametric tests. Other scholars instead recommend using multiple regression analysis controlling for any pre-treatment covariate that might have significantly affected the outcome (Green 2009, Gerber and Green 2012). We test the robustness of our results by using both approaches.

As we conclude this section, we emphasize that the decisions of the participants on the course requirements were binding on the entire class. We also highlight that the participants were unaware of the experiment and of the research project. The students believed only that they were participating in these discussions and the subsequent referenda to decide the requirements of their class (which was true – and our balanced experimental design did not impact the ultimate outcomes). Moreover, all scripted interventions were pedagogically justifiable. While different students were exposed to different arguments, none of the arguments were misleading. Different students were just prompted to think about different considerations, as is commonly done in small-group class discussions. We justified the presence of surveys as an evaluation tool to help the professor decide if the process should be replicated the subsequent year. Students are familiar with this type of survey because professors introducing innovation in the curriculum often conduct their own custom evaluation to obtain data mid-semester. In short, we made the setting feel completely natural.

As we turn to the results, it is noteworthy to mention the informal feedback that we received from our moderators. Treatment-moderators reported they felt paralyzed by the script and felt this limited them from having influence. We decided in advance that scripting the interventions of the moderators would limit the effect of particularly charismatic or persuasive moderators. We also felt that using a script would make clear the level of intervention, and make the experiment easy to explain and to replicate. Moreover, we wanted to see if even minor intervention could have an effect.

4. Model specification and results

To analyze the results of the experiment, we employ parametric and non-parametric methods that isolate each discussion in turn. We present the descriptive data and non-parametric tests in appendices 2 and 3, and focus in this section on the results of our regression analysis, presented in table 2. Columns 1 through 4 present the results from the first discussion (the writing requirement), and columns 5 through 8 refer to the second discussion (the exam requirement). As mentioned above, we place more weight on the first discussion, as the second may have been contaminated by the first. In fact, however, the main results hold across both discussions.

Columns 1 and 5 present the results from a linear probability model that estimates the probability that a student votes for option B in the post-discussion referendum. These baseline models include only the randomized treatment – there are no control variables. This clean approach is equivalent to a non-parametric estimate of the difference in means between the treatment effect in the treatment group and in the non-treatment group (see appendix 3).

Columns 2–4 and 6–8 present the results from various regression models. We first introduce a non-linear model (probit) and then introduce pre-treatment covariates as controls in the linear model. Some scholars argue against using a regression model that includes pre-treatment covariates to analyze experimental results because it may introduce bias (for an excellent discussion, see Freedman 2008a, 2008b). In some applications, however, the inclusion of pre-treatment covariates can improve the precision of the estimation of the treatment effect (Green

2009, Gerber and Green 2012). We present both approaches to test the robustness of our results.¹⁵

Columns 2 and 6 present the results of the non-linear specification (probit). We do not include pre-treatment covariates because our objective is merely to show that our results hold in the non-linear specification.¹⁶ Columns 3 and 7 present an ordinary least squares (OLS) specification that controls for the pre-discussion preferences of students.¹⁷ Columns 4 and 8 introduce pre-treatment covariates that control for specific characteristics of students: gender, graduation class, and membership in a varsity team (due to the constrained schedules of athletes, as mentioned above).¹⁸ These models further include indicators for the eight time slots during which the experiment was conducted and a variable that captures the percentage of students in each group in favor of option B.¹⁹ We include this last variable to ensure that the treatment effects are not driven by group pressure. Note that when we introduce the control variables, the interpretation of the coefficients differs subtly. In the baseline models (without controls), the coefficients indicate the effect of the treatment on the post-discussion preferences. In the models with controls for the pre-discussion preferences, the coefficients indicate the effect of the treatment on the *difference* between pre-discussion preferences and post-discussion preferences.

Note that the dependent variable in the second discussion is coded differently from the coding employed in the first discussion due to the presence of three indifferent students. It is coded zero if the student favors option A or is indifferent; it is coded one if the student favors option B. Thus the coefficient of treatment A shows the effect of treatment A on reducing the probability of the student voting for choice B. The effect is robust to switching the way we code the three indifferent students.

4.1 Results from the first discussion (writing requirement)

As table 2 shows, the effect of a moderator influencing toward option A in the first discussion is significantly different from zero (at the ten percent level) only in the linear and non-

¹⁵ All results hold when we bootstrap to simulate standard errors, and the standard errors of our treatments decrease. Results also hold when we employ a robust estimator for the variance in the linear probability model. These results are available in the replication materials.

¹⁶ Our main findings also hold in the probit models when we include the pre-discussion preferences. These results are available in the replication materials.

¹⁷ We introduce an indicator variable that assumes a value of one if the student's pre-discussion preference was option A, and a value of 0 otherwise. We introduce a second indicator variable that assumes a value of one if the student's pre-discussion preference was option B, and a value of zero otherwise. Our baseline category is pre-discussion indifference.

¹⁸ Membership on a varsity team is the only pre-treatment covariate that predicts treatment assignment, although not in all models. See appendix 2. Note that we control for "graduation class" (2008, 2009, 2010, or 2011) because we are concerned to control for the participant's years of experience as a student.

¹⁹ The indicators for section-meetings are introduced to control for potential differences across the students enrolled in the eight section-meetings. Given that students were randomly assigned to groups within their section-meeting, there is no need to adjust the standard errors of the regression for clustering. Clusters are randomly formed in this design.

linear baseline models (columns 1 and 2). The effect of a moderator influencing toward option B is statistically significant across all specifications at the five percent level or stronger.

Importantly, the majority of students held a pre-discussion preference for option A in the first discussion. Thus the imprecise estimate of moderator-influence toward option A is not surprising. We simply do not have enough data to isolate the effect of the treatment A when considering the difference from pre-discussion preferences.

So, the analysis of the first discussion shows a statistically significant moderator effect with regard to treatment B that is robust to a variety of different specifications. The effect of treatment A, which is constrained by the distribution of pre-discussion preferences in our sample, is weakly statistically significant only in the baseline models.

Interestingly, but maybe not surprisingly, the presence of a strict rule of discussion, which allows all participants to express an opinion, favors a change in opinion toward the less popular option B, independent of moderator influence. Thus a strict rule of discussion seems to have had a role in protecting the minority position – at least in this first discussion.

4.2 Results from the second discussion (exam requirement)

The second part of table 2 (columns 5 through 8) presents similar results for the second discussion. In the control group, a large majority of students (97 percent) held pre-discussion preferences in favor of option B, and therefore we observe that moderator influence toward option B is not statistically significant. Moderator influence toward option A is statistically significant at the ten percent level in the probit model and at the one percent level in all of the linear specifications. The discussion rule in the second discussion has no statistically significant impact.

Table 2: The effect of moderator-influence								
Discussion:	First discussion (writing requirement)				Second discussion (exam requirement)			
Outcome:	Option A=0, option B=1 (no cases of indifference)				Option A or indifferent=0,option B=1			
Estimator:	OLS	PROBIT	OLS	OLS	OLS	PROBIT	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Moderator influencing toward option A	-0.14* (-1.69)	-0.08* (-1.65)	-0.07 (-1.04)	0.04 (0.47)	-0.19*** (-2.66)	-0.14* (-1.81)	-0.22*** (-3.24)	-0.36*** (-3.64)
Moderator influencing toward option B	0.21** (2.50)	0.18** (2.00)	0.23*** (3.40)	0.57*** (4.51)	0.05 (0.69)	0.02 (0.79)	0.05 (0.80)	0.15* (1.93)
Discussion rules indicator	0.17** (2.37)	0.17** (2.03)	0.17*** (3.16)	0.33*** (4.79)	-0.08 (-1.36)	-0.07 (-1.31)	-0.06 (-1.09)	-0.04 (-0.64)
Constant	0.09 (1.43)	Included	0.56*** (4.11)	-5.71 (-0.10)	0.97*** (20.83)	Included	0.80*** (9.06)	93.25 (1.61)
Pre-discussion preferences			Included	Included			Included	Included
Student characteristics				Included				Included
Number of students in the group in favor of option B				Included				Included
Section-meeting indicators				Included				Included
Number of observations	107	107	107	107	107	107	107	107

Notes: *** indicates significance at the 0.01 percent level, ** at the 0.05 percent level, and * at the 0.10 percent level. For the probit models, we present the marginal effect for a discrete change of each indicator from zero to one (holding other indicators to zero). The numbers in parentheses are t-statistics (columns 1, 3, 4, 5, 7 and 8) or z-statistics (columns 2 and 6). “Student characteristics” control variables include the graduation class of participant, an indicator for gender, and an indicator for membership on a varsity team.

5. Discussion of results and suggestions for future research

Our experiment builds on the findings of previous research, which suggests the importance of moderator-influence (Fulwider 2005, Humphreys *et al.* 2006). We find that the expressed opinions of moderators can have an effect on deliberative discussion. The magnitudes of our estimated effects suggest that moderator influence could be decisive in a deliberative discussion. If moderators express an opinion even in a limited manner, as in our experiment, they can have a statistically and substantively significant impact on deliberation outcomes.

Future research should address a number of issues that our experiment does not. For example, we estimate smaller treatment effects in the second discussion. Given our experimental set-up, the first discussion treatment effects are unbiased – the students had no anticipation of treatment. The treatment applied during the first discussion may, however, have influenced the second discussion. We thus flag the attenuated moderator effect in the second discussion as a question for future investigation.

Our results also indicate smaller treatment effects when influencing towards the pre-discussion majority preference. This result may simply be driven by the constraints of the sample, as discussed above. Still, our experiment is not designed to identify differential effects between influencing toward majority and minority opinions. An experiment designed to vary the majority opinion in the groups could address such a question. Still, our regressions (column 4 and 8) do show that our results are robust to the inclusion of controls capturing the effect of group pressure. When isolating the effect of moderators from group pressure, the magnitude is greater (see table 2). The result suggests that a moderator might be particularly effective in defending minority opinions from group pressure.

Another result of our experiment worth further consideration is the effect of a strict rule of discussion – one that explicitly invites each participant to speak once before any other participant can speak twice. In this first discussion, the rule appears to promote the minority opinion, but the finding does not hold in the second discussion. Our discussion rules were not re-randomized for the second discussion because it would have appeared unnatural to change procedures between discussions. Thus we did not investigate this result experimentally. One possible explanation is that the number of students in the minority in the second discussion was so small that the rule of discussion did not have an effect.²⁰

Finally, we investigate the interaction of moderator–gender and student–gender. Surprisingly, the interaction is not statistically significant. Karpowitz, Mendelberg, and Shaker (2012) find an effect of gender group-composition on deliberation, so we expected the gender of the moderator to matter. Yet, we did not randomize the gender of the moderator or the group composition, so we flag this question for future research.

As for the external validity of our results, we highlight several issues to consider:

²⁰ Analyzing the interaction between the discussion rule and the moderator treatments, we find no statistically significant results in either discussion.

First, undergraduate students might be particularly susceptible to persuasion due to their young age. Some studies from the 1980s and 1990s, which examine attitude strength over an individual's lifecycle, show that age is inversely correlated with openness to changing opinion – although other studies challenge this finding.²¹ When we turn to laboratory experiments on the impact of framing, studies by Nelson, Clawson, and Oxley (1997) and Kühberger (1998) do not detect different effects across students and non-students. In their study of younger and older adults, Kim *et al.* (2006) find that senior citizens are actually more susceptible to the influence of framing than are undergraduate students. Note also that in laboratory experiments where participants must provide a rationale for their choices, the impact of framing diminishes.²² Of course, all these studies are conducted in a laboratory with subjects who are aware of being studied.²³ Thus, the relationship between age and malleability in a deliberative setting requiring justification is not obvious, and the results of the above laboratory experiments may not apply to our field experiment setting.

Second, if undergraduates perceive graduate students as experts, their preferences might be particularly malleable. Numerous studies in social psychology show that persuasiveness generally increases with communicator expertise and credibility.²⁴ Our experience at Yale University does not suggest that undergraduate students perceive moderators as figures of authority or as experts, but we cannot exclude this possibility. Still, in community settings, moderators are often local leaders, officials, or representatives of donors, who clearly do have some direct power over the participants – or they come from non-governmental organizations and have real expertise. In contrast, none of our moderators was serving (or had ever served) as the teaching assistant for anyone in his/her group, so the moderators had no grading power over their students. Our moderators had no experience in devising class requirements, and the information contained in their persuasive statements was common knowledge. The scripts were built using information provided to all students by the professor in class. Overall, while it is possible that undergraduate students perceived graduate students as experts, we believe that in many community settings the relative perception of the authority and expertise of the moderators is higher. Ultimately, however, this empirical question cannot be tested without replicating our experiment in a community deliberative decision-making process.

Third, our participants had well-defined *a priori* preferences over issues that they cared about and understood thoroughly (course requirements). In real-world applications of deliberative democracy, the knowledge and motivation of participants vary. In participatory budgeting, for example, only a small minority of participants understands the budgeting process. Numerous studies on persuasion in social psychology have shown that highly-motivated subjects who

²¹ Sears (1986), Krosnick and Alwin (1989), and Alwin, Cohen, and Newcomb (1991) have found a relationship; Tyler and Schuller (1991) and Visser and Krosnick (1998) have not.

²² See, for example, Takemura (1993), Sieck and Yates (1997), and Druckman (2004).

²³ Regarding the differences between field and laboratory experiments, see Levitt and List (2006, 2007a, 2007b, 2009), Falk and Heckman (2009), Kessler (2010), and Camerer (2011).

²⁴ See, for example, French and Raven (1960), Rhine and Severance (1970), Eagly and Chaiken (1993), Priester and Petty (2003), and Cialdini and Goldstein (2004). Also see Park (2012).

understand the topic of discussion are less prone to persuasion.²⁵ If this caveat holds, our findings likely understate the impact of moderator-influence.

Fourth, we deliberately constrained the ways in which our moderators could intervene in discussions (limiting them to three scripted statements). In community settings, such restrictions are absent. As noted in the introduction, from February to April 2009, one of the authors of this study participated in the weekly meetings of the Participatory Budgeting Council in Porto Alegre, Brazil (see Spada 2012). The author observed moderators who were not neutral and even used direct threats against participants. For example, they warned that participants who criticized the city government would not receive city funding for their projects.

While our experiment does not directly address the above issues and thus questions of external validity remain, our observations of real-world applications of deliberative democracy in community settings lead us to suspect that our study underestimates the effect of non-neutral moderators in many real-world applications of deliberative democracy.

6. Implications for the theory and practice of deliberative democracy

Endowed with the coercive power to intervene in discussions, moderators play a crucial role in deliberative democracy. Idealized as the neutral arbitrators of deliberations, theorists and practitioners have often ignored the potential misuse of moderator power. The results of this study suggest, however, that we may wish to rethink the role of moderators, lest we confound the effect of deliberation with the effect of moderator influence in real-world applications of deliberative democracy. While our study is certainly not designed to test various means of controlling moderators, the estimated impact of moderator influence does indicate a need to consider the issue.

Our experiment also provides a methodological contribution to the pragmatic research agenda on deliberative democracy (Fung 2007). Researchers can apply our fairly simple experimental design to analyze many elements of a real deliberative decision-making process. We suggest extensions of our work, from analyzing more in detail the effect of the provision of information within the discussion, to the effect of group size and composition, the length of the discussion, and, indeed, all aspects of deliberative decision-making processes.

From the standpoint of practitioners, we note that communities around the world are increasingly turning to deliberative and participatory governance processes, where moderators engage people in direct policymaking. The approach has been advocated by the World Bank, the International Monetary Fund, the United Nations, many non-governmental organizations, and municipal governments in developed and developing countries all over the world. According to estimates by Mansuri and Rao (2004, 2007), World Bank lending for “community driven

²⁵ See Petty and Caccioppo (1986), Zuwerink and Devine (1996), Petty and Wegener (1998), and Zuwerink Jacks and Devine (2000).

development” projects based on deliberative and participatory processes grew from US\$325 million in 1996 to US\$7 billion in 2010.²⁶

Scholars have warned us that participatory and deliberative approaches may be vulnerable to capture by local elites (see Platteau and Gaspart 2003, Baierle 2007, Wampler 2007, and Spada 2012). The specific ways in which elites may hijack these processes, however, have not been fully explored. One potential mechanism runs through the influence of moderators. As shown by our experiment, moderators might serve as an effective but subtle tool – one that is difficult to detect.

Even in cases where there is no manipulative intent by an interest group, the recruitment methods, backgrounds, and training of moderators might all play a role in deliberative processes. Often, for example, moderators are recruited from local civil society organizations, which deal with specific social issues (for example, education). Such knowledgeable and partisan moderators might *unintentionally* bias discussions towards their preferred outcomes. Our scripted interventions suggest that even a seemingly minor suggestion can impact the preferences of participants.

Paying careful attention to potential moderator influence may prove particularly useful for the study of the epistemic benefits of deliberative discussions and also for studies that analyze changes in attitudes, polarization, and the formation of consensus. At present, a puzzle exists in the literature. Consider the fact that laboratory research in political psychology has found that group discussions have a polarizing effect on the attitudes of participants (Sunstein 2000). In contrast, Deliberative Polls, and other similarly structured deliberations, find that group discussions do not lead to polarization (Luskin, Fishkin, and Hahn 2007). Note that the laboratory experiments do not employ moderators, while the Deliberative Polls do. Perhaps the difference in outcomes – polarization in un-moderated laboratory discussions and non-polarization in moderated Deliberative Polls – is indeed a moderator-driven effect. Deliberative Polls do employ professional moderators who are *instructed* to be neutral and not force consensus. Yet, many professional moderators might have preferences favoring consensus over polarization due to their extensive training in conflict resolution and their well-intentioned desire to provide an enjoyable experience for the participants. Such preferences might unintentionally bias the discussion against polarization. In his critical analysis of a European Deliberative Poll, Steiner (2012: 273) observes that paying the expenses of participants may further encourage a pleasant atmosphere. He concludes that a good atmosphere, reinforced by moderators who intervene in discussions to promote deliberation – plus the dissemination of pre-discussion information packages - may all work together to induce changes in attitudes and preferences. From an epistemic perspective, it is important to distinguish the effects of each of these various aspects of the experience from the inherent effects of the deliberation itself. Our research here specifically suggests greater scrutiny of the behavior of moderators in studies that aim to measure the effect of deliberation on attitude change. Researchers should seek to identify the separate effects of group discussions and moderator influence.

²⁶ For experimental work on community-driven reconstruction in the post-conflict setting of Liberia, see Fearon, Humphreys, and Weinstein (2009).

Our study thus aims at reducing the distances between the theory of deliberative democracy, the empirical testing of its consequences, and the actual practice of deliberative democracy in the real world. With this in mind, we conclude by highlighting simple policy recommendations that practitioners can consider to minimize the risk of manipulation by moderators and thereby strengthen deliberative processes.

First, we recommend that organizers of deliberative processes simply make their participants aware of the potential for moderator influence. The literature on persuasion shows that forewarning people of a speaker's persuasive intent can promote resistance (see Freedman and Sears 1965, Hass and Grady 1975, and Petty and Caccioppo 1979). Highlighting the possibility of misusing the moderator position may put the conveners and participants engaged in a deliberative procedure on guard against intentional and unintentional manipulation.

Second, several studies have shown that participants in discussions can learn resistance to suggestive messages (McGuire and Papageorgis 1961, Pfau *et al.* 1990, Compton and Pfau 2004, Banas and Reins 2010). Following this literature, deliberative processes might include capacity-building exercises that expose participants to weakly persuasive messages. Exposing participants to weak messages – such that they do not actually change attitudes – can teach participants to resist stronger forms of persuasion. This approach, of course, could make consensus-building more difficult, so we raise it with caution.

Third, we suggest assessing the preferences of moderators before conducting the deliberative procedure. In some settings, moderator influence may not be an issue. Where the moderators have little stake in the outcome of a deliberation, implementing complex moderation schemes might be unnecessary and actually prove disruptive. Where moderators have high stakes in the outcome, however, sterilizing the potential impact of moderators should take priority. Conveners could employ multiple moderators who represent different points of view and work together to balance their respective biases. We know from recent literature in political psychology that the effects of opposing framing messages can offset each other (Chong and Druckman 2013).

Finally, we recommend the random selection and assignment of moderators, a procedure that was effectively employed in São Tomé and Príncipe (see Humphreys *et al.* 2006). The random assignment of moderators may limit the ability of external interest groups to target them, and biases amongst the moderators may counterbalance each other. As an added benefit to scholars, increasing the use of random assignment would expand the prospects of conducting more field experiments on the role of moderator influence.

Appendix 1: Detailed description of the experiment

The experiment proceeded as follows. After breaking into groups, the students were given a survey. We began by soliciting pre-discussion preferences, as shown in table A1. The second page of the survey read, “Please do not turn the page until the moderator instructs you to do so. Thank you.”

Table A1: Pre-discussion survey

ON THIS PAGE ARE PRE-DISCUSSION SURVEY QUESTIONS.
THEY DO *NOT* COUNT FOR THE FINAL VOTE.
YOU WILL BE ALLOWED TO CHANGE YOUR MIND DURING THE DISCUSSION.

Select your preferred writing requirement option with an x:

- Three short papers (maximum 3 pages each)
- One short paper (maximum 2 pages) and one long paper (8 pages)
- I am indifferent

Select your preferred exam requirement with an x:

- One cumulative final
- One midterm and one non-cumulative final
- I am indifferent

The discussion over the writing requirement ensued. Two different discussion rules were randomly assigned to groups (irrespective of moderator “preference” assignment). According to one rule, discussion was left completely open. Specifically, students were instructed by the moderator, “*All students can speak as they like. Please be polite and listen to one another.*” According to the second rule, discussion was regulated. Specifically, students were instructed, “*Each student must be given a chance to speak before anyone can speak a second time. To ensure that everyone gets a chance, a student can talk for a maximum of 2 minutes for his/her first turn, and 1 minute for his/her second and third turn. This rule will be strictly enforced and I apologize in advance if I have to cut anyone off. But I will if necessary. I thank you in advance for your cooperation and understanding.*”

Next, moderators read one of the scripts from table A2, either expressing a preference or not. The students were then allowed to discuss the writing requirement. When the discussion time reached the halfway point, the moderators continued with the scripts in table A3. Student discussion was allowed to continue until the time limit was reached. When time was up, the three moderators concluded with the final intervention described in table A4. Students then voted on the writing requirement by secret ballot. They could choose between the two options above or choose indifference. After voting, the deliberation turned to the exam requirement. The experiment for this discussion followed the exact same format as the writing requirement discussion. “Preferences” of the moderators were randomly assigned (ahead of time).²⁷ The

²⁷ Thus, some placebo moderators from the first discussion were randomly assigned to be influencing moderators for the second discussion (and vice versa). We instructed such moderators to say s/he did not have any preference over the second decision if asked. No one asked.

discussion rule that had been randomly assigned for the writing requirement deliberation was maintained for the exam requirement. We omit the scripts for the second discussion; they are available on request.

Table A2: First discussion, opening intervention

Type of moderator	Script
Placebo	<i>As the TA has said, we have to deliberate first on the decision regarding the writing requirement. You have to decide between 3 short papers, 3 pages each, and 1 short paper and 1 long paper, 2 pages and 8 pages.</i>
In favor of three short papers	<i>Before letting you discuss, I want to share with you one little tip from my experience regarding this type of assignment. I take my job as moderator seriously and have notes: Don't be fooled by the fact that three assignments are more than two. Take into consideration the fact that writing a short paper might take less time than writing a long one. You might be able to write two good 3-page papers in less time than an 8-page paper takes.</i>
In favor of one long paper	<i>Before letting you discuss, I want to share with you one little tip from my experience regarding this type of assignment. I take my job as moderator seriously and have notes: Don't be fooled by the fact that writing a short paper requires less time than writing a long one. Writing three different short papers requires coming up with three different ideas. Coming up with ideas is the hardest part of this assignment. If you do one short and one long paper, Prof. Vreeland will allow you to use the short paper as a draft for the long paper. This means that in the end, if you choose the option of one short and one long paper, you simply have to come up with ideas for one paper and then just write one longer paper using the same ideas.</i>

Table A3: First discussion, midway intervention

Type of moderator	Script
Placebo	<i>Nothing</i>
In favor of three short papers	<i>At this point of the discussion I want to focus your attention on one important element: the preparation for the exam. Writing three short papers on three different topics means that you have already thoroughly prepared three major topics that will be covered in exam questions. This will give you more mastery of the subject and might help make studying for exam questions easier, regardless of the structure of exams.</i>
In favor of one long paper	<i>At this point of the discussion I want to focus your attention on one important element: the preparation for the exam. Writing three short superficial papers on three different topics means that you haven't mastered any single topic. On the other hand writing a long paper allows you to focus and master a topic. Thus, when the time of the exam comes, you have already prepared in depth some of the material for the final exam. This will give you more mastery of the subject.</i>

Table A4: First discussion, final Intervention

Type of moderator	Script
Placebo	<i>The time is almost up. It's time to stop the discussion. You have raised excellent points to support both proposals.</i>
In favor of three short papers	<i>Before voting I want to (point out)/(remind you of)²⁸ one final element: the importance of learning parsimony. When you compete for grants or make presentations, you will discover that often you are allowed to write only a couple of pages. Thus, from the point of view of pure usefulness for your future, the 3 short page papers is probably best.</i>
In favor of one long paper	<i>Before voting I want to (point out)/(remind you of) one final element: the importance of learning how to write a draft and then learning how to expand this draft into a proper project. This will be the process you will follow to write your senior essay and any type of other project in the future. Thus, from the point of view of pure usefulness for your future, one short paper and one long paper is probably best.</i>

²⁸ Moderators were allowed to say “remind you” instead of “point out” if this point was already raised by students. We did not want the scripts to seem awkward.

Appendix 2: Pre-discussion preferences and randomization checks

Table A5 (below) presents the pre-treatment preferences of the students and their discussion treatment assignment. As discussed in the main text, the majority of the students in the first discussion favored option A (80.3 percent), while a similar majority of students in the second discussion favored option B (79.4 percent). Note that there were five indifferent students in the first discussion and twelve in the second.

Regarding the randomization checks, we use Fisher's exact test to investigate the relationship between treatment assignment and pre-treatment preferences, as well as treatment assignment and the other covariates – see table A5. (Fisher's exact test is appropriate because we have multiple cells with an expected frequency of five or fewer.) The null hypothesis of the test we present in table A5 is that treatment assignment is independent of pre-discussion preferences. The analysis reassuringly reveals non-significant results for each discussion. Analysis with respect to the other individual characteristics that we consider – gender, age, and varsity team-membership – produces similar non-significant results. Finally, employing a multinomial logit to construct a joint chi-square test of the statistical significance of the pre-treatment covariates and treatment assignment indicates that none has a statistically significant relationship (results not presented but available in the replication materials). Again, we obtain similar results when analyzing the discussion rule, which we randomly assigned to some groups (results available upon request). Overall, our randomization procedure did not generate significant imbalances in treatment assignment. The multinomial logit does show, however, that membership in a varsity team weakly predicts treatment assignment. Thus, in table 2, we include a multiple regression model that controls for this variable.

Table A5: Pre-discussion preferences and randomization checks

Pre-treatment preferences in the first discussion (writing requirement)				
	In favor of A	In favor of B	Indifferent	Total
Treatment toward option A	30	4	0	34
Treatment toward option B	25	4	3	32
No influence	31	8	2	41
Total	86 (80.4%)	16 (14.9%)	5 (4.7%)	107 (100%)

Fisher's exact test: 0.377

Pre-treatment preferences in the second discussion (exam requirement)				
	In favor of A	In favor of B	Indifferent	Total
Treatment toward option A	1	24	2	27
Treatment toward option B	4	23	4	31
No influence	5	38	6	49
Total	10 (9.3%)	85 (79.4%)	12 (11.2%)	107 (100%)

Fisher's exact test: 0.725

Appendix 3: Post-discussion data, average treatment effects, and confidence intervals

The post-deliberation preferences for the first discussion (writing requirement) are as follows.

A3.1 First discussion (writing requirement):

In the control group (where moderators exerted no influence), 83 percent (34/41) of the students voted for option A (3 short response papers), while 17 percent (7/41) of the students voted for option B (1 short response paper and 1 final paper); no student was indifferent.

In groups where moderators influenced toward option A, 97 percent (33/34) of the students voted in favor of option A, while 3 percent (1/34) of the students voted in favor of option B; no student was indifferent. The average treatment effect appears to increase the number of students in favor of option A (or decreases the number of those in favor of option B) by 14 percentage points. A difference-in-means test indicates that this result is statistically significant (the 95 percent confidence interval suggests that the effect is between 0.05 percent and 28 percent).

In groups where the moderators influenced toward option B, 62.5 percent (20/32) of the students voted in favor of option A, while 37.5 percent (12/32) of the students voted in favor of option B; no student was indifferent. The average treatment effect appears to increase the number of students in favor of option B (or decreases the number of students in favor of option A) by 20.5 percentage points. A difference-in-means test indicates that this result is statistically significant (the 95 percent confidence interval suggests that the effect is between 0.07 percent and 41 percent).

A3.2 Second discussion (exam requirement):

Similar results emerge from the second discussion, but the interpretation is a bit more complex due to the presence of students who declared indifference. In the control group 6.1 percent (3/49) of the students voted in the referendum for option A (cumulative final exam), and 93.9 percent (46/49) voted for option B (midterm examination and non-cumulative final examination); no student was indifferent.

In groups where moderators influenced toward option A, 18.5 percent (5/27) of the students voted for option A, 7.4 percent (2/27) were indifferent, and 74.1 percent (20/27) voted for option B. The treatment appears to have reduced the number of students in favor of option B by 19.8 percent. This result is driven by (1) an increase in the number of students in favor of option A (by 12.4 percentage points) and (2) an increase in the number of indifferent students (by 7.4 percent). A difference-in-means test indicates that the result is statistically significant (the 95 percent confidence interval suggests that the effect is between 4 percent and 36 percent).

In groups where moderators influenced toward option B, no student voted for option A, 3 percent (1/31) of the students were indifferent, and 97 percent (30/31) voted in favor of option B.

Thus the treatment reduced the number of students in favor of option A by 6 percentage points. This result is driven by (1) an increase in the number of students in favor of option B (by 3 percentage points), and (2) an increase in the number of indifferent students (by 3 percentage points). A test of the difference in means shows, however, that this result is not statistically significant (the 95 percent confidence interval suggests that the effect is between -3 percent and 15 percent). Again, the non-result is partly due to the fact that an overwhelming majority preferred this outcome to begin with (79.4 percent – see table A5).

References

- Ackerman, B. 1991. *We the People: Foundations*. Cambridge. Harvard University Press.
- Ackerman, B. 1998. *We the People: Transformations*. Cambridge. Harvard University Press.
- Alwin, D. F., Cohen, R. L., and Newcomb, T. M. 1991. *Political Attitudes over the Lifespan: The Bennington Women after Fifty Years*. Madison. University of Wisconsin Press.
- Austen-Smith, D., and Feddersen, T. 2006. Deliberation, Preference Uncertainty and Voting Rules. *American Political Science Review* 100, 2, 209-217
- Bächtiger, A., Niemeyer, S., Neblo, M., Steenbergen, M., and Steiner, J. 2010. Disentangling Diversity in Deliberative Democracy: Competing Theories, their Blind Spots and Complementarities. *Journal of Political Philosophy* 18, 1, 32-63.
- Baierle, S. G. 2007. *Lutas Urbanas em Porto Alegre: Entre a Revolução Política e o Transformismo*. Porto Alegre. Cidade – Centro de Assessoria e Estudos Urbanos.
- Banas, J. A., and Rains, S. A. 2010. A Meta-analysis of Research on Inoculation Theory. *Communication Monographs* 77, 281-311.
- Barber, B. R. 1984. *Strong Democracy*. Berkeley. University of California Press.
- Benhabib, S. 1992. *Situating the Self: Gender, Community and Postmodernism in Contemporary Ethics*. New York. Routledge.
- Benhabib, S. 1996. *Democracy and Difference: Contesting the Boundaries of the Political*. Princeton. Princeton University Press.
- Besson, S., and Martí, L. 2006. *Deliberative Democracy and Its Discontents*. Aldershot. Ashgate.
- Bohman, J. 1996. *Public deliberation: Pluralism, Complexity and Democracy*. Cambridge. MIT Press.
- Bohman, J. 1998. Survey Article: The Coming of Age of Deliberative Democracy. *The Journal of Political Philosophy* 6, 4, 400-425.
- Bohman, J. 2009. Epistemic Value and Deliberative Democracy. *The Good Society* 18, 2, 28-34.
- Boulle, L. 2011. *Mediation: Principles Process Practice*. Sydney. Lexis Nexis.
- Camerer, C. F. 2011. The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List (December 30, 2011). Available at SSRN: <http://ssrn.com/abstract=1977749>.
- Chambers, S. 1996. *Reasonable Democracy: Jurgen Habermas and the Politics of Discourse*. Ithaca. Cornell University Press.
- Chambers, S. 2003. Deliberative Democratic Theory. *Annual Review of Political Science* 6, 307-326.
- Chong, D., and Druckman, J. N. 2013. Counter-framing Effects. *The Journal of Politics* 75, 1, 1-16.

- Cialdini, R. B., and Goldstein, N. J. 2004. Social Influence: Compliance and Conformity. *Annual Review of Psychology* 55, 591-621.
- Cialdini, R. B. 2007. *Influence: The Psychology of Persuasion*. New York. Harper Collins Publishers.
- Cohen, J. 1986. An Epistemic Conception of Democracy. *Ethics* 97, 26-38.
- Cohen, J. 1989. Deliberation and Democratic Legitimacy. In Hamlin, A., and Pettit, P. *The Good Polity*. Oxford. Blackwell.
- Cohen, J. 1996. Procedure and Substance in Deliberative Democracy in Benhabib, S. *Democracy and Difference: Contesting the Boundaries of the Political*. Princeton. Princeton University Press.
- Compton, J., and Pfau, M. 2004. Use of Inoculation to Foster Resistance to Credit Card Marketing Targeting College Students. *Journal of Applied Communication Research* 32, 343-364.
- Dickson, E., Hafer, C., and Landa, D. 2008a Cognition and Strategy: A Deliberation Experiment. *Journal of Politics* 704, 974-89.
- Dickson, E., Hafer, C., and Landa, D. 2008b Rules of Debate: Theory and Experiment. Typescript. New York University.
- Druckman, J. N. 2004. Political Preference Formation. *American Political Science Review* 98, 671-686.
- Dryzek, J. 2000. *Deliberative Democracy and Beyond: Liberals, Critics and Contestations*. Oxford. Oxford University Press.
- Dryzek, J. 2010. *Foundations and Frontiers of Deliberative Governance*. Oxford. Oxford University Press.
- Eagly, A. H., and Chaiken, S. 1993. *The Psychology of Attitudes*. New York. Harcourt Brace Jovanovich College Publishers.
- Elster, J. 1997. The Market and the Forum: Three Varieties of Political Theory. In Bohman, J., and Rehg, W. *Deliberative Democracy: Essays on Reason and Politics*. Cambridge. MIT Press.
- Elster, J. 1998. *Deliberative Democracy*. New York. Cambridge University Press.
- Elstub, S. 2010. The Third Generation of Deliberative Democracy. *Political Studies Review* 83, 3, 291-307.
- Estlund, D. 1997. Beyond Fairness and Deliberation: The Epistemic Dimension of Democratic Authority. In Bohman, J., and Rehg, W. *Deliberative Democracy: Essays on Reason and Politics*. Cambridge. MIT Press.
- Falk, A., and Heckman, J. J. 2009. Lab Experiments are a Major Source of Knowledge in the Social Sciences. *Science* 326, 535-538.

- Farrar, C., Green, D. P., Green, J. E., Nickerson, D., and Shewfelt, S. 2009. Does Discussion Group Composition Affect Policy Preferences? Results from Three Randomized Experiments. *Political Psychology* 304, 615-647.
- Farrar, C., Fishkin, J. S., Green, D. P., List, C., Luskin, R. C., and Paluck, E. L. 2010. Disaggregating Deliberations Effects: An Experiment within a Deliberative Poll. *British Journal of Political Science* 40, 2, 333-347.
- Fearon, J., Humphreys, M., and Weinstein, J. M. 2009. Can Development Aid Contribute to Social Cohesion After Civil War? Evidence from a Field Experiment in Post-Conflict Liberia. *American Economic Review: Papers and Proceedings* 992, 287-291.
- Fishkin, J. 1995. *The Voice of the People*. New Haven. Yale University Press.
- Fishkin, J. 2009. *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford. Oxford University Press.
- Freedman, J. L., and Sears, D. O. 1965. Warning, Distraction, and Resistance to Influence. *Journal of Personality and Social Psychology* 1, 262-266.
- Freedman, D. 2008a. On Regression Adjustments to Experimental Data. *Advances in Applied Mathematics* 40, 180-193.
- Freedman, D. 2008b. On Regression Adjustments in Experiments with Several Treatments. *Annals of Applied Statistics* 2, 176-96.
- French, J. P. R., Jr., and Raven, B. 1960. The Bases of Social Power. In Cartwright, D. and Zander, A. *Group Dynamics*. New York. Harper and Row.
- Fulwider, J. 2005. Do Moderators Matter? Answering a Jury Deliberation Challenge to Deliberative Democracy. Paper prepared for the annual meeting of the American Political Science Association, Washington, DC.
- Fung, A., and Wright, E. 2003. *Deepening Democracy: Experiments in Empowered Participatory Governance*. Princeton. Princeton University Press.
- Fung, A. 2007. Democratic Theory and Political Science: A Pragmatic Method of Constructive Engagement. *American Political Science Review* 101, 3, 443-458.
- Gastil, J. 1993. *Democracy in Small Groups: Participation, Decision-making and Communication*. Philadelphia. New Society.
- Gastil, J. 2000. *By Popular Demand: Revitalising Representative Democracy Through Deliberative elections*. Berkeley. University of California Press.
- Gaventa, J. 2006. Deepening the Deepening Democracy Debate. IDS Working Papers.
- Gerber, A., and Green, D. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York. W. W. Norton and Company.
- Goodin, R. 2005. Sequencing Deliberative Moments. *Acta Politica* 40, 182-196.
- Goodin, R. 2008. *Innovating Democracy*. Oxford. Oxford University Press.

- Green, D. 2009. Regression Adjustments to Experimental Data: Do David Freedman's Concerns Apply to Political Science? Paper prepared for the 26th Annual Society for Political Methodology Summer Conference, Yale University.
- Gutmann, A., and Thompson, D. 1996. *Democracy and Disagreement*. Cambridge. Harvard University Press.
- Gutmann, A., and Thompson, D. 2004. *Why Deliberative Democracy*. Princeton. Princeton University Press.
- Habermas, J. 1984. *Theory of Communicative Action, vol. 1: Reason and the Rationalization of Society*. (translated by Thomas McCarthy). Boston. Beacon Press.
- Habermas, J. 1987. *The Theory of Communicative Action, vol. 2*. (translated by Thomas McCarthy). Boston. Beacon Press.
- Habermas, J. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge. MIT Press.
- Hafer, C., and Landa D. 2005. Deliberation and Social Polarization. Manuscript. New York University.
- Hafer, C., and Landa D. 2007. Deliberation as Self-Discovery. *Journal of Theoretical Politics* 193, 329-60.
- Hass, R. G., and Grady, K. 1975. Temporal Delay, Type of Forewarning, and Resistance to Influence. *Journal of Experimental Social Psychology* 11, 459-469.
- Humphreys, M., Masters, W., and Sandbu, M. 2006. The Role of Leaders in Democratic Deliberations: Results from a Field Experiment in São Tomé and Príncipe. *World Politics* 58, 583-622.
- Imai, K., and Yamamoto, T. 2010. Causal Inference with Differential Measurement Error: Nonparametric Identification and Sensitivity Analyses. *American Journal of Political Science*, 542, 543-560.
- Karpowitz, C., and Mansbridge, J. 2005. Disagreement and Consensus: The Importance of Dynamic Updating in Public Deliberation. In Gastil, J. and Levine, P. *The Deliberative Democracy Handbook: Strategies for Effective Civic Engagement in the Twenty-First Century*. San Francisco. Jossey-Bass.
- Karpowitz, C., Mendelberg, T., and Shaker, L. 2012. Gender Inequality in Deliberative Participation. *American Political Science Review* 106, 3, 533-547.
- Kessler, J. B., 2010. Strategic and Informational Environment: Addressing the Lab-field Debate with Laboratory Gift Exchange Experiments. Working Paper.
- Kim, S., Goldstein, D., Hasher L., and Zacks, R. 2006. Framing Effects in Younger and Older Adults. *The Journal of Gerontology* 60, 215-218.
- Krosnick, J. A., and Alwin, D. E. 1989. Aging and Susceptibility to Attitude Change. *Journal of Personality and Social Psychology* 57, 416-425.

- Kühberger, A. 1998. The Influence of Framing on Risky Decisions: A Meta-analysis. *Organizational Behavior and Human Decision Processes* 75, 1, 23-55
- Landa, D. 2005. The Moral Economy of Deliberative Participation. Manuscript. New York University.
- Landa, D., and Meirowitz, A. 2009. Game Theory, Information, and Deliberative Democracy. *American Journal of Political Science* 53, 2, 427-444.
- Landmore, H. 2012. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton. Princeton University Press.
- Levine, P., Fong, A., and Gastil J. 2005. Future Directions for Public Deliberation. *Journal of Public Deliberation* 11, 1-13.
- Levitt, S. D., and List, J. A., 2006. What do Laboratory Experiments tell us About the Real World? Working Paper.
- Levitt, S. D., and List, J. A., 2007a. What do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *The Journal of Economic Perspectives* 21, 153-174.
- Levitt, S. D., and List, J. A., 2007b. Viewpoint: On the Generalizability of Lab Behaviour to the Field. *Canadian Journal of Economics* 40, 347-370.
- Levitt, S. D., and List, J. A., 2009. Field Experiments in Economics: The past, the Present, and the Future. *European Economic Review* 53, 1-18.
- List, C., Luskin, R. C., Fishkin, J. S., and McLean, I. 2006. Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls. PSPE Working Paper.
- Luskin, R. C., Fishkin, J. S., and Hahn, K. S. 2007. Consensus and Polarization in Small Group Deliberations. Paper prepared for the annual meeting of the American Political Science Association, Chicago, IL.
- Mansbridge, J. 1980. *Beyond Adversarial Democracy*. Chicago. University of Chicago Press.
- Mansbridge, J. 1999. Everyday Talk in the Deliberative System. In Macedo, S. *Deliberative politics: Essays on Democracy and Disagreement*. New York. Oxford University Press.
- Mansbridge, J., Bohman, J., Chambers, S., Estlund, D., Follesdai, A., Fung, A., Lafont, C., Manin, B., and Marti, J. L. 2010. The Place of Self-interest and Role of Power in Deliberative Democracy. *The Journal of Political Philosophy* 18, 1, 64-100.
- Mansbridge, J., Bohman, J., Chambers, S., Christiano, T., Fung, A., Parkinson, J., Thompson, D. F., and Warren, M. E. 2013. A Systemic Approach to Deliberative Democracy. In Parkinson, J. and Mansbridge, J. *The Deliberative System*. New York. Cambridge University Press.
- Mansuri, G., and Rao, V. 2004. Community-Based and -Driven Development: A Critical Review. *The World Bank Research Observer* 19, 1, 1-39.
- Mansuri, G., and Rao, V. 2007. Update Note on Community-Based and -Driven Development. Mimeo. Washington DC. World Bank.

- McGuire, W. J., and Papageorgis, D. 1961. The Relative Efficacy of Various Types of Prior Belief-defense in Producing Immunity Against Persuasion. *Journal of Abnormal Social Psychology* 62, 327-337.
- Meirowitz, A. 2007. In Defense of Exclusionary Deliberation: Communication and Voting with Private Beliefs and Values. *Journal of Theoretical Politics* 193, 301-27.
- Mendelberg, T. 2002. The Deliberative Citizen: Theory and Evidence. *Political Decision Making, Deliberation and Participation* 6, 151-193.
- Morrison, K., and Singer, M. M. 2007. Inequality and Deliberative Development: Revisiting Bolivia Experience with the PRSP. *Development Policy Review* 256, 721-740.
- Muhlberger, P. 2007. Strategic and Communicative Rationality in a Deliberative Field Experiment. Paper prepared for the annual meeting of the Midwest Political Science Association, Chicago, IL.
- Mutz, D. C. 2008. Is Deliberative Democracy a Falsifiable Theory? *Annual Review of Political Science* 11, 521-38.
- Mutz, D. C., and Pemantle, R. 2011. The Perils of Randomization Checks in the Analysis of Experiments. Paper prepared for the annual meeting of the Society for Political Methodology.
- Myers, C. D. 2010. Information Sharing in Small Group Deliberation. Manuscript. Princeton University.
- Nelson, T. E., Clawson, R. A., and Oxley, Z. M. 1997. Media Framing of a Civil Liberties Conflict and its Effect on Tolerance. *American Political Science Review* 91, 567-583.
- Nino, C. S. 1996. *The Constitution of Deliberative Democracy*. New Haven: Yale University
- Page, S. E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools and Societies*. Princeton. Princeton University Press.
- Park, J. Y. 2012. Testing Conditional Effects of a Moderator in Deliberation: A Lab Experiment. Paper prepared for the annual meeting of the American Political Science Association, New Orleans, LA.
- Pateman, C. 1970. *Participation and Democratic Theory*. New York. Cambridge University Press
- Petty, R. E., and Cacioppo, J. T. 1979. Effects of Forewarning of Persuasive Intent and Involvement on Cognitive Responses and Persuasion. *Personality and Social Psychology Bulletin* 5,173-176.
- Petty, R. E., and Cacioppo, J. T. 1986. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York. Springer.
- Petty, R. E., and Wegener, D. T. 1998. Attitude Change: Multiple Roles for Persuasion Variables. In Gilbert, D., Fiske, S., and Lindzey, G. *The Handbook of Social Psychology*. New York. McGraw-Hill.

- Pfau, M., Kenski, H. C., Nitz, M., and Sorenson, J. 1990. Efficacy of Inoculation Strategies in Promoting Resistance to Political Attack Messages: Application to Direct Mail. *Communication Monographs* 57, 25-43.
- Pierce, J. L., Neeley G., and Budziak, J. 2008. Can Deliberative Democracy Work in Hierarchical Organizations? *Journal of Public Deliberation* 4, 1, 14.
- Platteau, J., and Gaspard, F. 2003. The Risk of Resource Misappropriation in Community-Driven Development. *World Development* 31, 10, 1687-1703.
- Priester, R. J., and Petty, E. R. 2003. The Influence of Spokesperson Trustworthiness on Message Elaboration, Attitude Strength, and Advertising Effectiveness. *Journal of Consumer Psychology* 13, 408-21.
- Rao, V., and Mansuri, G. 2012. *Localizing Development: Does Participation Work?* Washington. World Bank.
- Rhine R. J, and Severance, L. J. 1970. Ego-involvement, discrepancy, source credibility, and attitude change. *Journal of Personality and Social Psychology* 16, 2, 175-90.
- Sanders, L. M. 1997. Against deliberation. *Political Theory*, 253, 347-376.
- Sears, D. O. 1986. College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature. *Journal of Personality and Social Psychology* 51, 515-530.
- Shapiro, I. 1999. Enough of Deliberation: Politics is About Interests and Power. In Macedo, S. *Deliberative Politics: Essays on Democracy and Disagreement*. New York. Oxford University Press.
- Shapiro, I. 2002. Optimal deliberation? *Journal of Political Philosophy*, 10, 2, 196-211
- Shapiro, I. 2003. *The State of Democratic Theory*. Princeton. Princeton University Press.
- Sieck, W., and Yates, J. F. 1997. Exposition effects on decision making. *Organizational Behavior and Human Decision Processes* 70, 207-219.
- Spada, P. 2012. Political Competition in Deliberative and Participatory Institutions. Dissertation manuscript, Yale University.
- Steiner, J. 2012. *The Foundation of Deliberative Democracy: Empirical Research and Normative Implications*. New York. Cambridge University Press.
- Sunstein, C. R. 2000. Deliberative Trouble? Why Groups Go to Extremes. *The Yale Law Journal* 1101, 71-119.
- Takemura, K. 1993. The Effect of Decision Frame and Decision Justification on Risky Choice. *Japanese Psychological Research* 35, 36-40.
- Tyler, T. R., and Schuller, R. A. 1991. Aging and Attitude Change. *Journal of Personality and Social Psychology*, 61, 689-697.
- Visser, P. S., and Krosnick, J. A. 1998. Development of Attitude Strength over the Life Cycle: Surge and Decline. *Journal of Personality and Social Psychology* 75, 1389-1410.

- Wampler, B. 2007. *Participatory Budgeting in Brazil: Contestation, Cooperation, and Accountability*. University Park. Pennsylvania State University Press.
- Warren, M. 1992. Democratic Theory and Self-transformation. *American Political Science Review* 86, 8-23.
- Warren, M., and Pearse, H. 2008. *Designing Deliberative Democracy: the British Columbia Citizens Assembly*. New York. Cambridge University Press.
- Winters, M. S. 2007. Explaining Selection into Indonesias Kecamatan Development Program. Paper prepared for the annual meeting of the Midwest Political Science Association, Chicago, IL.
- Young, I. 1996. Communication and the Other: Beyond Deliberative Democracy. In Benhabib S. *Democracy and Difference: Contesting the Boundaries of the Political*. Princeton. Princeton University Press.
- Young, I. 1999. Justice, Inclusion and Deliberative Democracy. In Macedo S., *Deliberative Politics: Essays on Democracy and Disagreement*. Oxford. Oxford University Press.
- Young, I. 2000. *Inclusion and Democracy*. Oxford. Oxford University Press.
- Young, I. 2001. Activist Challenges to Deliberative Democracy. *Political Theory* 29, 5, 670-690.
- Zuwerink, J. R., and Devine, P. G. 1996. Attitude Importance and Resistance to Persuasion: It's not just the Thought that Counts. *Journal of Personality and Social Psychology* 70, 931-944.
- Zuwerink Jacks, J. R., and Devine, P. G. 2000. Attitude Importance, Forewarning of Message Content, and Resistance to Persuasion. *Basic and Applied Social Psychology* 22, 19-29.