

Social Choice in the Wild

Deliberative decision making and group decision stability.

(Preliminary draft -- comments welcomed.)

Peter Towbin (ptowbin@ucsc.edu)
UCSC Economics Department, 03/03/2014

Kenneth Arrow's impossibility theorem in social choice theory is remarkable for being at once both definitive and inscrutable. It seems to disallow a rational procedure for collective choice without suggesting practical ways to reduce the probability or magnitude of the problem. To shed light on this issue, and considering the difficulties of recreating real-world political dynamics in the lab, we study a pressing social dilemma in a coastal California city. Local citizens engaged in small group deliberations over strategies to resolve an intensifying water supply shortage -- an issue which had provoked sharply divided public opinion. We propose a measure of decision stability to grade the danger of a social choice paradox and find that the process of small group deliberation significantly increases this measure of stability.

Social choice theory and the democratic process

Individual rationality plays a central role in economic theory. The paradigm of utility maximization leads to conclusions about behavior only to the extent that agents are actually capable of understanding and consistently taking action to achieve preferred outcomes. One of the standard assumptions of what constitutes rationality is transitivity of preference: if an individual prefers option A over option B and prefers B over option C, then they necessarily prefer A over C. Social choice theory examines whether this notion of rationality can be scaled up to decision making by groups. Majority voting is one method groups might use to make a "social choice", but the theory considers all possible decision making procedures, aside from the dictatorial case of always following a particular individual's preference.

Given that a group's members are individually rational, is there a way to aggregate their preferences while guaranteeing that group decisions meet the same standard of rationality assumed for its individual members? Starting from seemingly very reasonable assumptions, the startling answer provided by Kenneth Arrow was that it is not [Arrow, 1953]. Any method used for aggregating a "social choice" is prone to violating the very basic characteristic of rational judgment described above, transitivity of preference. An important strand of conservative political theory drew on this result to argue for minimal government [Riker, 1982]. Given that the theorems of welfare economics insure that market-based solutions at least achieve Pareto optimal allocations, it was argued that even the most well intentioned democratic system would be capricious and manipulable, in contrast. Some saw this as a potentially fundamental challenge to the legitimacy of democratic governance [List et al, 2006].

On the opposite end of the spectrum are those that see this as motivation for deeper citizen engagement through deliberative democracy. Rather than a passive process of aggregating intact preferences, group dialog and democratic engagement is seen as the opportunity to transform preferences based on shared insight [Miller, 1992]. As deliberative and participatory approaches to governance have entered the

mainstream [Jacobs, Cook and Carpin, 2009], interest in the more arcane relation to social choice theory has also grown [Dryzek and List, 2003]. Ultimately, this is an empirical question.

Experiments in "Deliberative Polling"

James Fishkin has championed the concept of a deliberative opinion poll as a nimble corollary to deliberative democracy [Fishkin, 1991]. Rather than requiring deep public engagement on every issue, governing bodies can gain insight into public preferences on complicated and contentious issues by having a randomized sample of constituents participate in a deliberative dialog. Their post-deliberation preferences attempt to provide a representative sample of what public opinion *would be* if there was sufficient opportunity and time for more extensive participation. Farrar, et al, [2010] have provided some of the first empirical evidence on social choice by deliberative process. To understand their result, first let us look at a potential escape-hatch from the impossibility theorem. One of the assumptions of Arrow's theorem is that any decision making procedure should operate successfully no matter what the "input", as given by the profile of individual preferences that comprise the group. This is sometimes referred to as the assumption of "impartial culture", and implies that no a-priori structure can be assumed to exist in the probability distributions of the population's preferences. In fact, Duncan Black first pointed out one kind of structure that eliminates the paradox in his study of committee voting, called "single peakedness" [Black, 1948].

Imagine laying out, along a line, the set of alternatives that are being considered. For a given individual, imagine placing their most favored alternative down first. Then, on one side you place the second most valued alternatives and on the other side place the third alternative, alternating sides until every alternative has been placed. For this individual there is a single preference peak, and in either direction, preferences for the options decline. Now imagine adding another individual. If the line is not single-peaked for the second individual (the peak may be at a different location than the other individual, but still declining in both directions), it is possible that a reordering of the options can be found such that the line is single peaked for both individuals. This line is referred to as the structuring dimension. If there is a structuring dimension that is single peaked for all individuals, then their will be a (non-dictatorial) decision making procedure that preserves group preference transitivity.

Farrar, et al and List, et al [2013] argue that deliberation increases collective structure in the distribution of preferences, such that groups move closer to single-peakedness. Their measure is global, in the sense that it looks at changes across the entire population of deliberators. On the other hand, we don't learn what happens between individuals within the same deliberating group. An additional limitation to their measure is that while single-peakedness is a sufficient condition for transitive group preferences, it is not a necessary condition. The method proposed in this paper addresses some of these concerns.

Water, public good and debating desalination

One of the challenges of modern democracy is the sheer complexity of some of our societal choices. As we have overcome some of the past threats to our well being through science and technology, inevitably, the remaining or developing threats contain additional elements that render a techno-fix more problematic. Deliberative engagement may be useful, but potentially may require new approaches for 1) issues that require a conceptually demanding analysis, and 2) issues that are driven largely by emotion. The topic of this study demanded both complex information assimilation as well as processing of emotional content: should residents of Santa Cruz, Ca. build a costly sea water desalination plant to augment the water supply?

Residents were concerned by the cost of the plant as well as its energy consumption and other potential environmental impacts. But they had to weigh this against declining aquifers threatened with contamination from encroaching seawater. There was also increasing vulnerability to drought, tightened federal restrictions on accessing rivers and streams in an effort to resuscitate native fish species, and complex hydrogeological considerations that had already launched more than one PhD dissertation.

Though water is a necessity, only a fraction of what we use is a critical necessity and the rest supports a lifestyle and consumptive choices that Americans are accustomed to making freely. But water from desalination is about ten times more expensive than existing water sources, with the cost of building the plant well over \$100 million. Due to the cost and increasingly entrenched viewpoints from public authorities and groups advocating for or against the plant, desalination also proved to be an emotionally charged issue.

Presenting the different perspectives

After studying the water supply and desalination issue, reading local news coverage and responses left by readers in the newspapers' online forums, and going to a number of debates between representatives of the pro and con perspectives, I partitioned the issue along five dimensions. These represented five independent lines of argument that had been made for or against building the plant: two in favor, two opposed, and one that was positioned as neutral which urged further research into legal and regional factors bearing on the proposal. I developed a ten page Issue Guide for participants to review and discuss in a small group setting. The five perspectives formed the architecture for introducing the most salient facts and pertinent values that had been argued in the actual ongoing debate. I fact-checked the points with published research or the public record (water department research reports, newspaper accounts of problems at established desalination plants, etc) and had the document reviewed by members of the water agency, advocates of desalination, and leaders from anti-desalination organizations.

The key factual points were introduced (roughly) in equal proportion among the five sections, with each section arguing its case from the values laden perspective of an advocate of that view. In this way, readers had exposure to both core facts and the most salient values that had arisen in the literature and debates on the topic. During small group meetings, we took turns with participants reading a paragraph or so of the issue guide and then taking a moment to discuss or clarify what was covered before rotating to the next person. The design of the reference material in this fashion was intended to mitigate common problems that have been noted in group discussion.

In the natural flow of group discussion participants look at each fact with respect to the ultimate question at hand -- whether to build the desalination plant, in this case. Typically, this draws individuals into expressing their opinion and arguing their perspective from the beginning, before even the basic facts have been laid out. Biased information processing, the preferential acquisition of information that supports one's preconceived beliefs, is particularly strong when one anticipates having to defend one's self [Mojzisch, 2008]. Another documented issue with deliberation is that small groups are not likely to have a balanced representation of diverse views, or equally important, an even balance of argumentative charisma. The tendency for groups with an initial disposition in favor of one view is to follow what has been dubbed the "law of group polarization" [Sunstein, 2002] and become even more disposed towards the original bias. For these reasons, the group was instructed to focus the discussion on the five positions presented in the guide. They were not forbidden from expressing their own views (and inevitably they did) but the balance of time was spent looking at the strengths and

weaknesses of the constructed positions rather than staking out and defending their own views.

Survey on seawater desalination and representative values

An initial survey was conducted in person, with a team going door-to-door over several neighborhoods in the Santa Cruz and Soquel Creek water districts. Neighborhoods were chosen to sample the socio-economic range of the region. The response rate was close to 50% of households contacted, and this was the first self-selection challenge in getting a truly representative sample. The second and more substantial challenge was finding enough individuals from the initial survey group who were willing to join one or more deliberative sessions. Participants were paid roughly \$30 per hour for two hour small group meetings or \$100 for a three hour large group meeting.

The survey used in the door-to-door sample was given again, without modifications, when the participant arrived at their first group session and at the end of every group session. We elicited their level of support for the desalination plant itself as well as for a hypothetical list of citizen-candidates for a desalination task force. In this scenario, task force members would make a recommendation on the desalination plant after conducting an in-depth, independent review of the facts. For the initial survey, participants were only provided with a one sentence summary from each candidate, expressing that candidate's initial attitude about the plant (see the survey in Table 2, below). These attitudes had been selected to correspond to the most prevalent "camps" in the public debates. The survey ratings for different candidates were driven by the respondent's own views towards the candidate summary positions and also modeled the gradual revelation of candidate information that occurs in representative elections.

There were eight small group discussion sessions in which group members read aloud and discussed an Issue Guide which provided an overview of the water supply problem and then presented the five positions of the hypothetical candidates in more detail. Each position was motivated by a primary value orientation that was independent of the other perspectives. Thus, for example, it would be possible for a respondent to highly support the candidacy of all five prospective committee members, to indicate that they highly valued having each value orientation represented on the panel (this did occur but was atypical.) The body of argument for each candidate was roughly one page and contained core facts that pertained to the final decision, so that by the end of the five arguments participants had engaged with most of the central issues.

Survey Response and Group Attendance	
483	-- door-to-door surveys of town residents
37	-- survey respondents attend a small group
24	-- others recruited to small groups
61	-- total number of attendees in small groups
31	-- small group attendees also attend big group
18	-- survey respondents attend big group only
49	-- total number of attendees at big group

Table 1: Number of people and level of participation

Desalination Plant Survey

Name or ID _____

Survey # _____

The Santa Cruz and Soquel Creek water districts are considering building a seawater desalination plant to supplement our water supply. If a committee of citizens was established to judge whether or not to build the proposed desalination plant, what viewpoints would you be most comfortable with for the citizens sitting on that committee?

Each candidate seeking the committee position honestly explains their values and existing opinions about the desalination plant. But they also vow to take in new information by intensively studying the topic along with other committee members and revise their views in order to find the best solution for the community at large. If the following is all you hear from the candidate, how strongly would you favor or oppose them joining the committee:

Candidate 1:

It is important to maintain our quality of life and our local economic vitality.

(+) Initial attitude: favors the desalination plant.

Candidate 2:

We need to live within our means through conservation.

(-) Initial attitude: opposes the desalination plant.

Candidate 3:

First address the the bigger picture: agricultural water use and overly complex water laws.

(?) Initial attitude: can't decide on desalination until the bigger picture is better understood.

Candidate 4:

Diversify our water supply for a more sustainable and flexible water management.

(+) Initial attitude: favors the desalination plant.

Candidate 5. Consider global warming -- desalination is energy intensive.

(-) Initial attitude: opposes the desalination plant.

100 = Strongly Favor

50 = Don't care, either way.

0 = Strongly Oppose

Table 2: Description of candidates given in the survey.

There was also one large group that everyone was invited to which featured presentations from four experts -- two advocating in favor of the plant and two advocating in opposition. The participants then broke into smaller groups for discussion and to generate questions for the presenters. And finally, the large group was reconvened with an open discussion between presenters and attendees. The breakdown of how many participated in which part of the experiment is given in Table 1. (Unfortunately, scheduling issues prevented many of the attendees from joining both sessions and the results reported here are focused mostly on the initial small group discussions.)

Preferences for desalination before group deliberation

The question which is typically of primary interest in a deliberative polling event concerns what difference deliberation made on the final preferences for the actual topic of deliberation. A growing literature argues that deliberation has a substantive and salutary impact on policy preferences [Barabas, 2004; Luskin & Fishkin, 2002]. Making this case is not entirely straightforward since there is no objectively correct reference point for preference outcomes. The data from this experiment does provide some support for these arguments from the initial pattern of preference adjustment, rather than its final direction.

On the left side of Table 3 is a kernel density estimation followed by a boxplot of the distribution of levels of support for the desalination plant in the initial sample of all 483 households. While the median voter is almost exactly ambivalent, we see that the range of feelings span the gamut but there are distinct spikes of individuals with strong feelings against and in favor of the proposal.

The plots on the right reveal a quite striking result. For all those people who came to a small-group deliberation, these show their preference distributions from when they were surveyed door-to-door (on top) and the distribution from an identical survey when they arrived at their first deliberation (below). Note that there was no contact with or information provided to these participants between these two surveys. The top right plot looks quite similar to the overall initial distribution on its left. It is still bimodal but has a somewhat larger representation of the desalination skeptics. But the lower image shows that in the time between their initial off-the-cuff survey at home and when they arrived at their first group discussion, their preference distribution has substantially shifted -- towards neutrality.

This was unexpected and I can only speculate on plausible explanations. The first hypothesis considers that a large majority of the initial survey population was relatively uninformed on this issue. Even though it had been generating ongoing press coverage and community organizing, most people were familiar with only a small percentage of the relevant facts. Some of these people registered a neutral vote but most gave a gut reaction that was either yea or nay, depending on how they saw it fitting into their framework of values and political orientation. Once they made a commitment to attend a deliberative event to learn about and discuss the topic, they began to follow the issue more closely through discussions with friends and/or reading news accounts. Considering the level of complexity that is actually involved, by the time they arrived at the deliberation event they had an appreciation of their relative ignorance and could no longer fit the question neatly in an established corner of their political values framework. If this hypothesis is correct, the data reflects favorably on the flexibility of average citizens to reappraise their judgment and acknowledge humility. It also speaks to the inadequacy of standard telephone or in-person surveys which the deliberative polling approach is trying to remedy. On the other hand, when participants came to the deliberative event they knew that the goal was to discuss the issue with an open mind. Thus, an alternate explanation is that they simply reported their opinion to be more neutral than it actually was, because they felt that they were expected to be initially neutral. The observed effect is likely to be a combination of these factors.

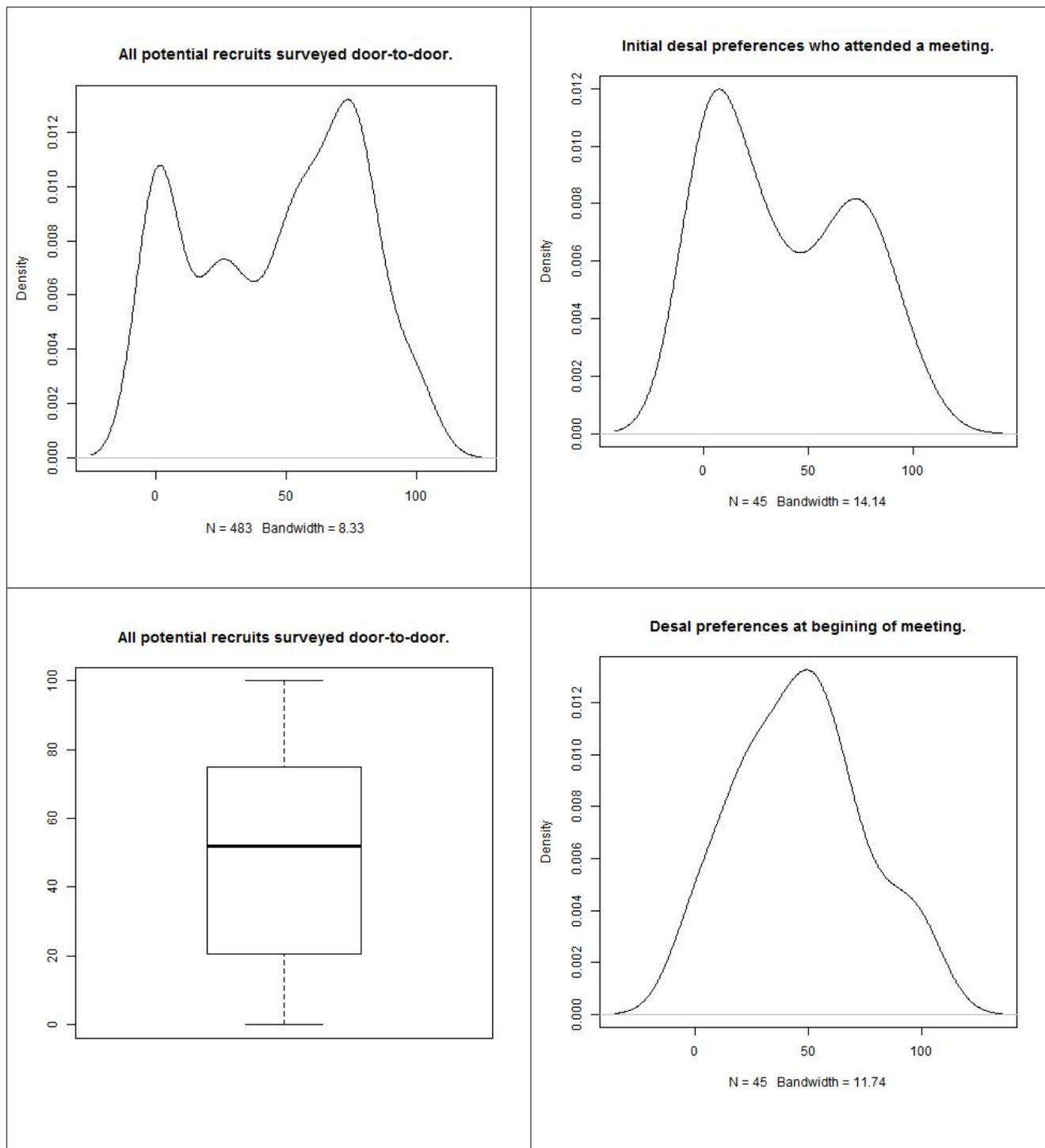


Table 3: Desal Preferences: Left: All door-to-door. Right: The ones who attended.

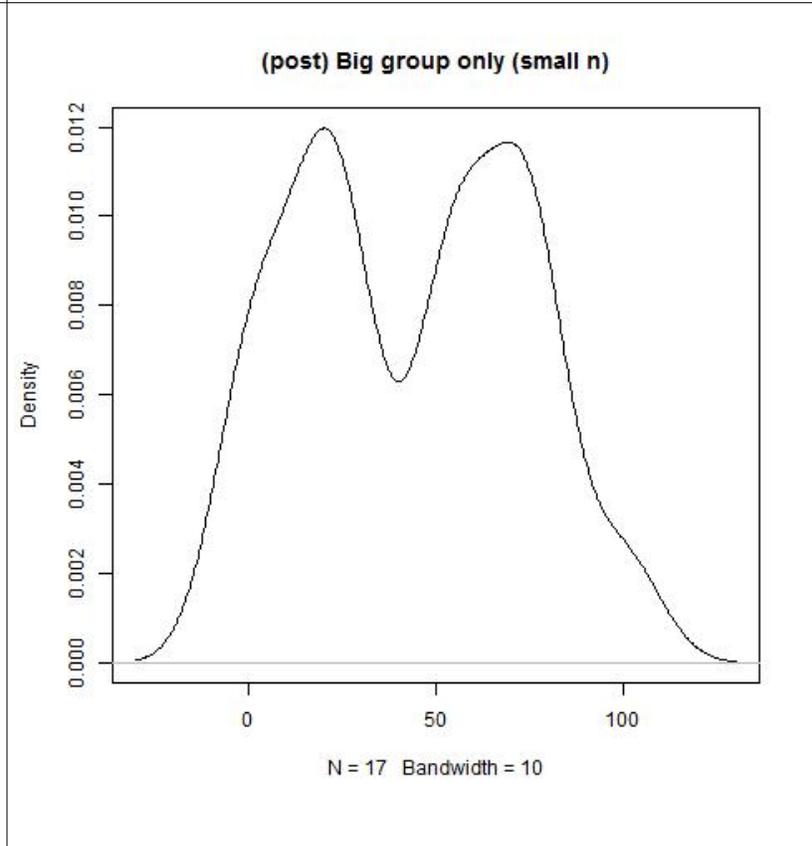
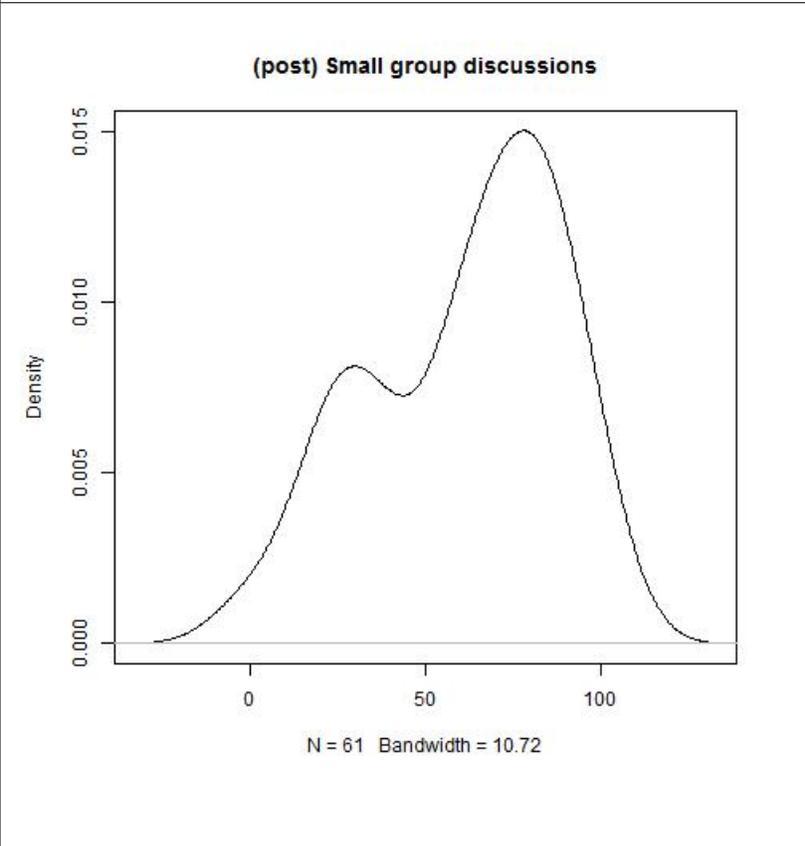
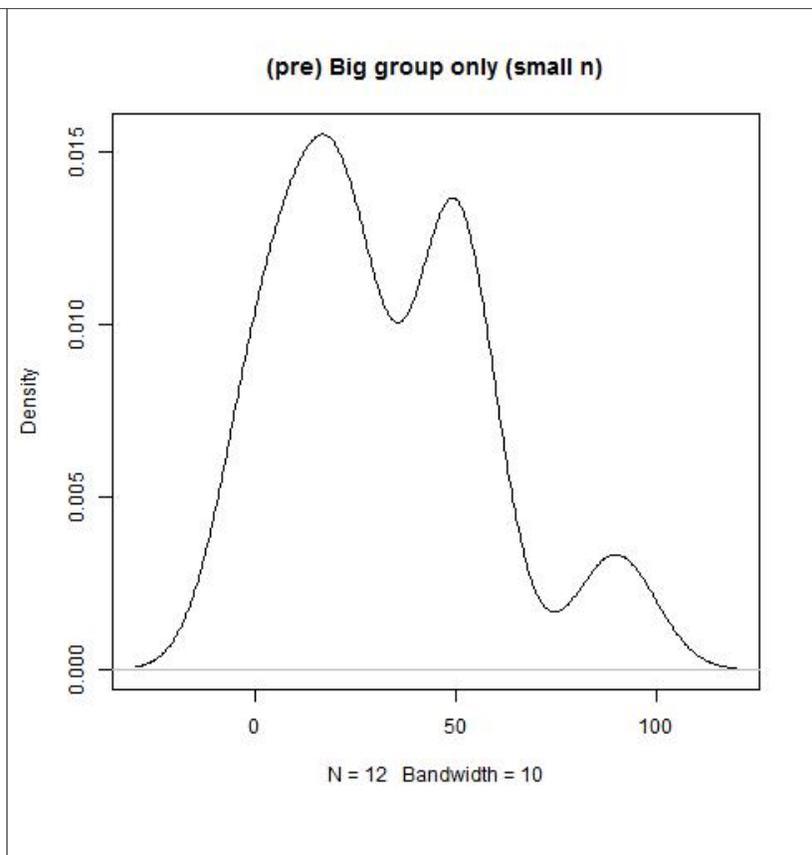
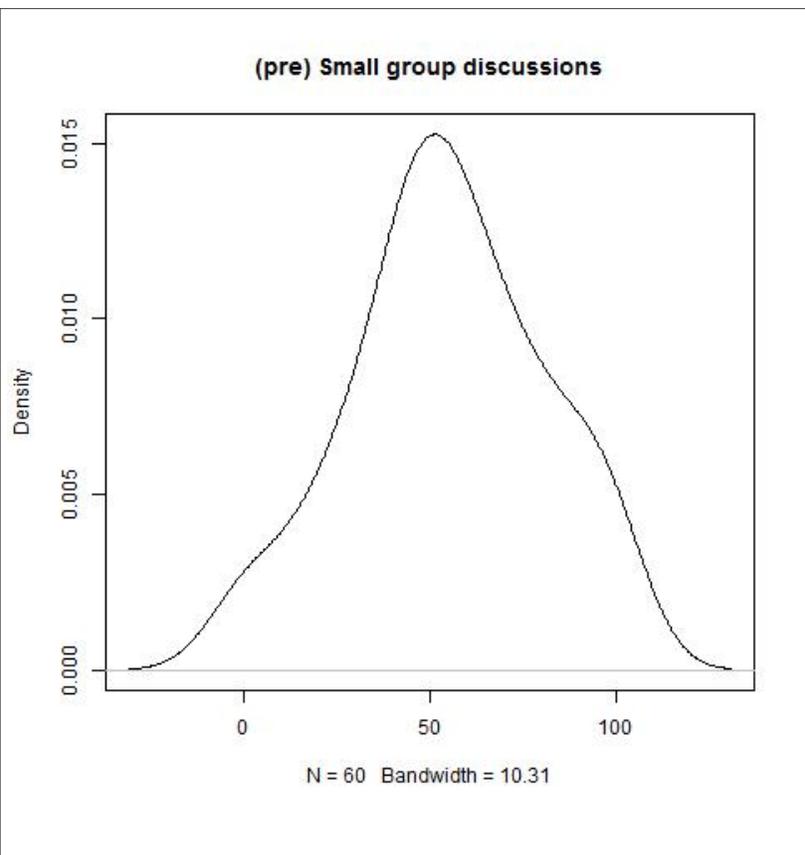


Table 4: Desal preference shift: small group (left) and big group only (right).

Impact of deliberation on preference for building a desalination plant

Table 4 shows the before and after density maps for individuals attending the small group (left side) and those individuals who only attended the big group (right side). Note that the before, or "pre", preferences shown here are from surveys at the beginning of the meetings, and the earlier door-to-door surveys are no longer considered. Also, recognize that the right hand side should be considered cautiously because of the relatively small number who only made it to the large group session and because a number of these failed to turn in the pre-deliberation survey. But the distributions on the right are consistent with the trend shown on the left, that as people learn more about the issue, they are more willing to make the investment and pay for the electricity required to build and operate the desalination plant.

Finally, Table 5 shows that the trend continues: those who attended both a small group meeting and a large group meeting have shifted the most in favor of the desal plant. The left side shows the density plot for this group and the right side shows the trend in box plots from 1)pre-small group, to 2)post-small group and then 3)post small and large groups.

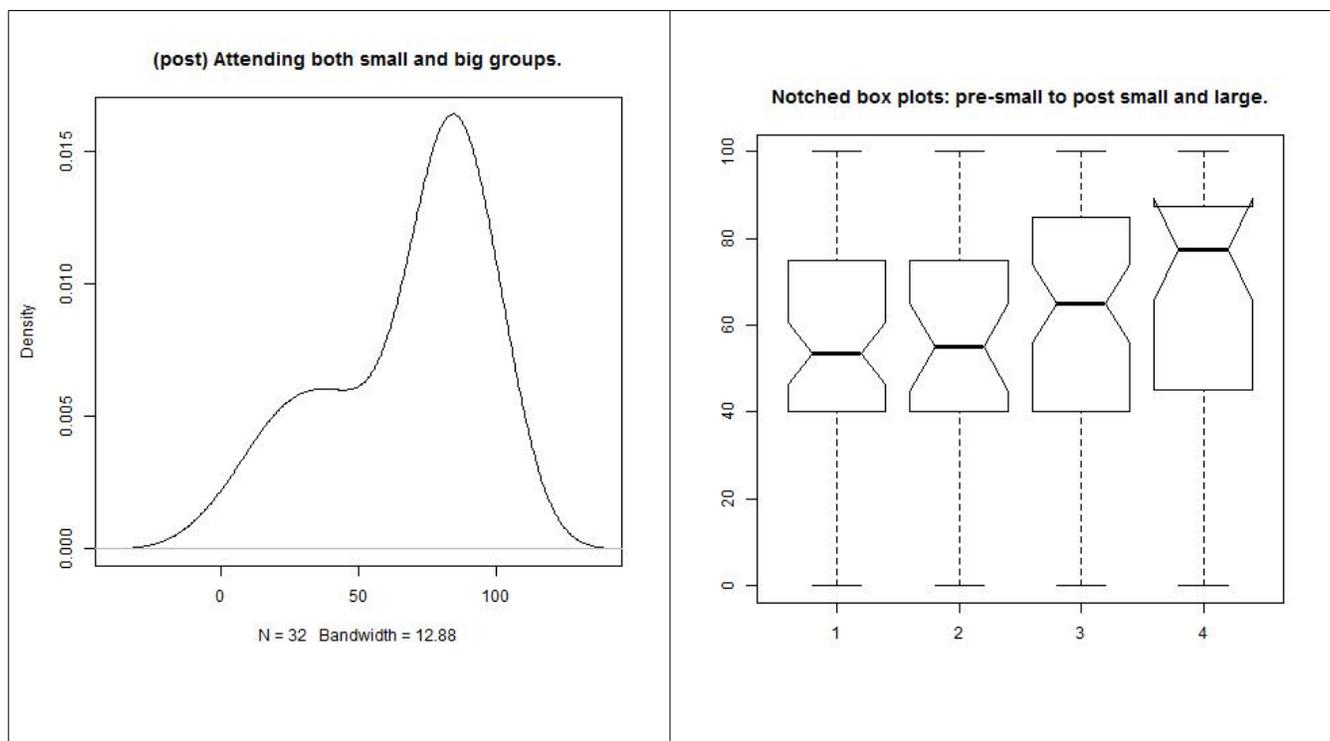


Table 5: Left: Desal preference density of those who went to both small and large. Right: 1)Pre-small group, 2) Pre-small who made it through big group as well. 3) Post-small group, 4) Post small and big for those who did both.

Structure out of chaos: an exploratory analysis

Although the shift in preferences is important, it is not the primary question addressed in this paper. Deliberative democracy theorists have argued that the dialog process itself may generate a shared structural correlation between the different dimensions of the preference space. The increasing correlation eliminates the viability of the "impartial culture" assumption of Arrow's theorem.

The first approach I took to assess the changing correlation structure is through linear regression. If correlations between different policy questions increases, the information-theoretic content of the preference profiles should become smaller and some policy issues should become more predictive of others. Given the breakdown of the survey into the five candidates plus the issue of the desalination plant, it was natural to look at whether preferences for the candidates become more predictive of preferences for the plant.

The first regression, in Table 6 below, looks at preferences from the initial door-to-door survey. Due to the relatively large number of respondents, we are able identify relatively weak correlations, with an R-squared of around .3. This indicates that there are pre-existing structural correlations among these components in the population at large. This is already at odds with a pure notion of "impartial culture".

Coefficients:				FORMULA: desal ~ c1 + c2 + c3 + c4 + c5
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	28.19875	3.90358	7.224	2.01e-12 ***
c1	0.25595	0.04123	6.208	1.17e-09 ***
c2	-0.11872	0.04329	-2.743	0.00633 **
c3	-0.03947	0.03709	-1.064	0.28777
c4	0.32432	0.03778	8.584	< 2e-16 ***
c5	-0.10302	0.04063	-2.535	0.01155 *

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.68 on 477 degrees of freedom
 Multiple R-squared: 0.3059, Adjusted R-squared: 0.2987
 F-statistic: 42.05 on 5 and 477 DF, p-value: < 2.2e-16

Table 6: (Door to door survey) Preferences regressed: desalination versus candidates.

	Estimate	Std. Error	t value	Pr(> t)	FORMULA: desal ~ c1 + c2 + c3 + c4 + c5
(Intercept)	40.82716	16.56830	2.464	0.01695 *	---
survey1\$c1	0.38119	0.14011	2.721	0.00875 **	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
survey1\$c2	-0.20704	0.15075	-1.373	0.17532	Residual standard error: 22.23 on 54 degrees of freedom
survey1\$c3	0.01498	0.13024	0.115	0.90884	Multiple R-squared: 0.3306, Adjusted R-squared: 0.2686
survey1\$c4	0.12073	0.14323	0.843	0.40301	F-statistic: 5.333 on 5 and 54 DF, p-value: 0.0004696
survey1\$c5	-0.07036	0.15111	-0.466	0.64335	

Table 7: (Pre small group) Preference for desalination regressed against candidate preferences.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	31.98502	11.90967	2.686	0.00955 **	FORMULA: desal ~ c1 + c2 + c3 + c4 + c5 --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 16.21 on 55 degrees of freedom Multiple R-squared: 0.6719, Adjusted R-squared: 0.6421 F-statistic: 22.52 on 5 and 55 DF, p-value: 3.214e-12
survey2\$c1	0.38319	0.11159	3.434	0.00114 **	
survey2\$c2	-0.27543	0.10864	-2.535	0.01411 *	
survey2\$c3	-0.02629	0.08633	-0.305	0.76187	
survey2\$c4	0.48169	0.10961	4.395	5.11e-05 ***	
survey2\$c5	-0.08414	0.10209	-0.824	0.41342	

Table 8: (Post small group) Preference for desalination regressed against candidate preferences.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	26.995871	23.211431	1.163	0.26742	FORMULA: desal ~ c1 + c2 + c3 + c4 + c5 --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 11.53 on 12 degrees of freedom Multiple R-squared: 0.9294, Adjusted R-squared: 0.9 F-statistic: 31.62 on 5 and 12 DF, p-value: 1.647e-06
survey7\$c1	0.470627	0.109678	4.291	0.00105 **	
survey7\$c2	-0.329328	0.243230	-1.354	0.20069	
survey7\$c3	-0.210380	0.122596	-1.716	0.11183	
survey7\$c4	0.639469	0.177087	3.611	0.00357 **	
survey7\$c5	0.006753	0.171126	0.039	0.96917	

Table 9: Regression after a reflective follow-up discussion, after small and large group events.

Table 7 shows the results from surveys just prior to small group discussions. As noted above, participants adjusted their opinions about desalination between their first encounters with the door-to-door survey and actually coming to a group discussion. But there is very little change with respect to overall predictive correlation with candidate opinions -- it looks very much like the results from the initial door-to-door survey. On the other hand, Table 8 shows the regression just after small group discussion. Attitudes toward conservation and, especially, towards the importance of having a more diverse set of choices for our water supply became statistically significant in predicting the preference for desalination. This relationship was significant in the large door-to-door sample. But the pre small group sample was too small to capture what was very diffuse knowledge about the status and sources of our water supply. This relationship became more broadly relevant to desalination preferences through the process of deliberation, after which the R-squared measure has risen above .6, indicating significant predictive power. From this data we can't tell if this was due to the impact of information provided in the Issue Guide and/or the exchange of ideas and feelings during deliberation, but some clues will emerge in the next section.

The final regression should be considered as only anecdotal evidence, but intriguing. This was a small follow up meeting (18 participants) for individuals who had already been to one or both of the small and large group events. I recruited individuals non-randomly in an attempt to get close to a balance between desalination supporters and opponents. The meeting was reflective in that discussion focused on what was driving the different opinions of pro, anti, and neutral or undecided participants. We tried to clarify what the different needs drove each group and found that the only substantial difference was in the need for information. The pro-desalination group was satisfied that alternative options had been fully explored and felt confident that desalination was the most effective solution. The three hours of discussion in this group seemed to me to generate the greatest insights into each other's opinions. This was my subjective interpretation of the evening and the R-squared score of .9 provides some additional support.

Decision stability and Condorcet winners

The final section looks at the social choice question directly through an analysis of Condorcet winners in the small groups that deliberated together as well as through a number of virtual comparisons to gain insight into these results. Given our field of five candidates, there are a number of reasonable procedures a group might use to select its choice of preferred candidate. For example, everyone could vote for their top pick and the candidate with the highest number of votes becomes the group's choice. But, with more than two candidates it is possible that one of the other candidates is actually preferred by a majority over the winner, leading to alternative procedures involving runoffs, Borda counts, etc. Moreover, having a majority winner does not eliminate the possibility that there are preference cycles in pairwise comparisons. Transitive group preferences are guaranteed in the case that there is a Condorcet winner.

Candidate X is considered a Condorcet winner if for every binary comparison between candidate X and one of the alternative candidates, Y, candidate X would win the majority of votes over Y. From our preference survey of each small group, we can easily calculate whether there would have been a Condorcet winner. In comparing X and Y, for each member of the group we assign a vote to X if and only if the member assigned a higher preference level for candidate X than candidate Y. Groups may fail to have a Condorcet winner due to preference cycling, but in small groups with even numbers of members it is also common to have tie votes. Thus, for small groups with random preferences (uniformly distributed from 0 to 100 for each candidate, in our case), the probability of a Condorcet winner is about 75% given an odd number of group members but only 30 to 40% for an even number of group members.

Group Size	Winning Candidate #	Least Margin
11	3	4
10	5	1
5	2	1
10	2	1
6	No Winner	0
7	3	1
5	No Winner	0
6	3	1
Sum of margins statistic =		9

Table 10: Condorcet results from pre survey for each of the eight small groups.

Table 10 shows the the results of a Condorcet contest for the preference profiles of each of the small groups, from the pre-meeting survey. The columns show the number of members in the group, which candidate was the Condorcet winner (or "No Winner" in some cases), and the "least margin". Least margin is the smallest difference in votes between the winner and the candidate(s) that came in second place. Clearly, the larger the least margin the more stable the outcome is to small deviations of preferences. At the bottom of the least margin column is the sum of all the winning margins, giving a summary statistic of the Condorcet stability of this 8-group set of preference profiles. Table 11 shows how things have changed in the post small group survey.

Group Size	Winning Candidate #	Least Margin
11	2	2
10	No Winner	0
5	2	3
10	2	3
6	4	2
8	4	2
5	4	2
6	No Winner	0
Sum of margins statistic =		14

Table 11: Condorcet results from post survey for each of the eight small groups.

From the sum of margins we can see that Condorcet stability is significantly greater after deliberation. Given the sum of margins statistic it is now straightforward to evaluate the hypothesis of an impartial culture. We can draw from uniform random preferences to simulate its distribution for this collection of group sizes. The graphics below depict this distribution and the lower image shows points in the distribution where the pre and post small group survey statistics lie.

The first image shows the histogram for all the resulting margin sums for a simulation with 150,000 samples of random preferences. A large number was used in order to get a reasonable estimate of the tail of the distribution. The second image shows a kernel density estimate for this distribution and the points in the distribution where the pre and post small group surveys lie. These can also be assessed quantitatively by looking at the proportions of runs that are as or more extreme than these values.

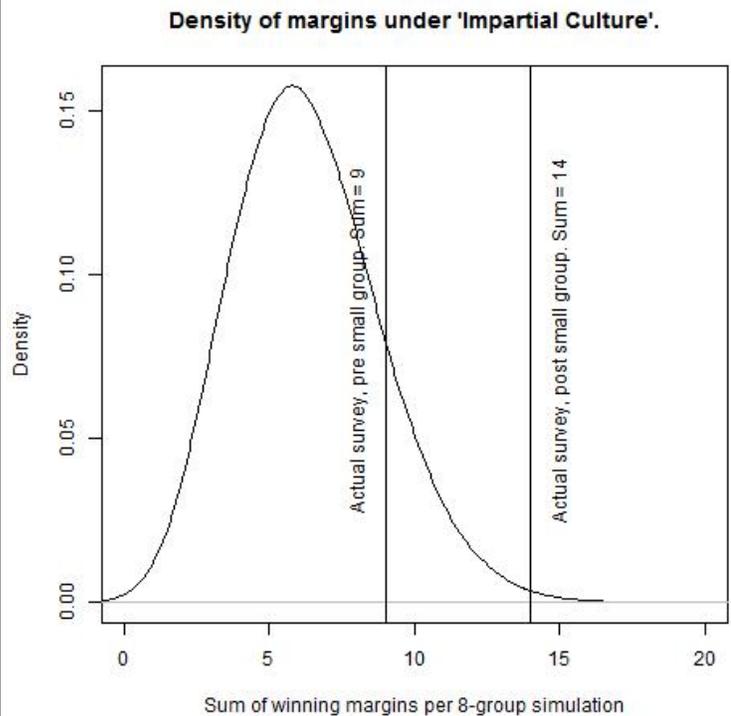
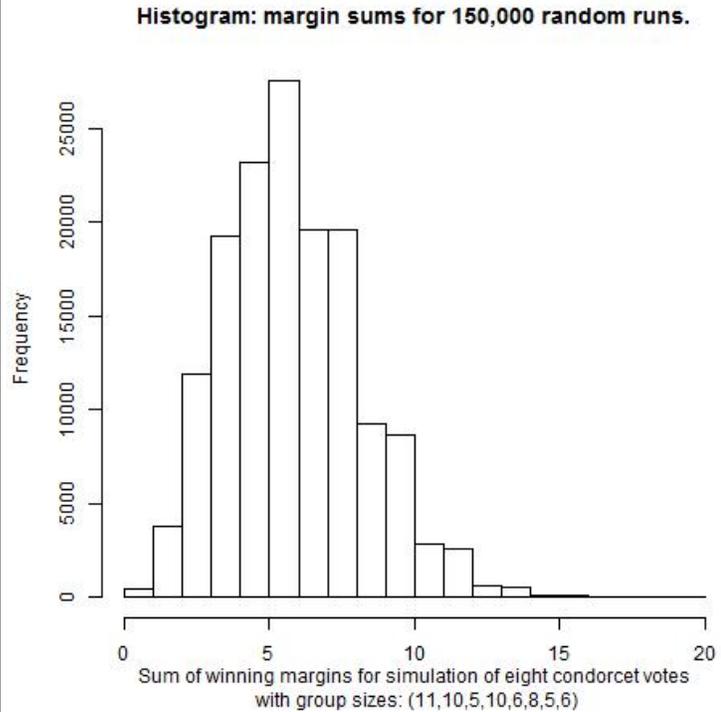
For the pre small group survey, the proportion of margin sums that are greater than or equal to 9 is 0.17. This means that we could reject the hypothesis of an impartial culture with 83% confidence. This is indicative but not conclusive.

However, after deliberating in a small group format we can conclusively reject the assumption of impartial culture: the probability of a margin sum value being greater than or equal to 14 is estimated as .0045 from the simulation. We can reject this hypothesis with greater than 99% confidence.

Given that our post-deliberation groups exhibit structure in their preference profiles that is not consistent with impartial culture, we would like to attribute the change to some aspect of the deliberative experience. This could be from reading the Issue Guide and collectively absorbing shared information, the exchange of ideas and comparison of perspectives allowed by the dialog process, or some other aspect. In order to evaluate this I performed another simulation. Let us take the collection of preferences of all small group participants at the pre-group survey as a representative sample, rather than using randomly generated preferences. Indeed, the groups were composed of one potential drawing from this sample distribution and the only thing determining which individuals fell together in a given group was their choice of a meeting time, which is plausibly unrelated to their preference distribution over desalination candidates.

Our new hypothesis to consider is that deliberation was not the cause of the jump in margin sums, rather it was a random outcome from a new distribution that was not qualitatively different from the incoming pre-group sample. We can evaluate this with a similar approach to the impartial culture simulation, but this time choosing the pre-group discussion sample distribution as representing the population, rather than a uniform random distribution. We then take draws from this population and assign them to groups of the same sizes as in our actual deliberation. Specifically, we sample without replacement to fill a particular group but then reset to the initial sample population before starting the next group. In the end, we calculate the margin sum of the hypothetical groups using the sampled preference profiles.

Results of this simulation show that the probability of getting a margin sum greater than or equal to 9 is 40%. This is quite reasonable considering we did get a margin sum of 9 in the actual draw. Also, the probability of getting greater than or equal to 14 is only 4.4%, so we reject this with 95% confidence. This is encouraging, but we still don't have a sense of what aspect of the deliberative process led to the increase in preference correlations. Perhaps we could have just emailed the Issue Guide to everyone and had the same effect on preference adjustment?



To evaluate this, we next take the collective set of preferences determined by the post-group survey and consider this as representative of the population distribution. Sampling from this (without replacement within a group, with replacement between groups) reveals that the probability of our statistic being greater than or equal to 14 is actually slightly less now, 3.0%. The implication of this, with the caveat that this is one small study, is that most (if not all) of the correlation structure that pushes these groups towards a transitive and more stable Condorcet winner is coming from the dialog with other group members. This is not to say that the information provided in the Issue Guide is irrelevant, indeed they are engaging in dialog about or stimulated by the Issue Guide. The point is that preferences have become correlated within the small groups themselves to increase Condorcet stability, while there is no indication that any broad, out of group, change in decision stability has developed. The face to face interaction through information-based small group deliberation was crucial for overcoming the impossibility result of rational social choice.

If this result can be reliably replicated, it would support the assertions of deliberation enthusiasts that small groups can achieve more coherent and rational decision making through deliberative methods -- at least to the extent that intransitive preferences materially contribute to poor group choices. Face to face deliberation was crucial in structuring collective preferences, while studying identical background information failed to show this effect. On the other hand, the fact that this measure was essentially unchanged when sampling from the post-deliberation population casts doubt on the implications of the result for large scale decision making, such as for elected officials or ballot initiatives. Or, this may be a function of the extreme complexity of the question addressed. Regression analysis supported the idea that, globally, preference relationships are becoming increasingly correlated. Water policy is impacted by a overwhelming number of complicated factors. Perhaps two hours of deliberation was insufficient to translate this global increase in cardinal preference correlations into a global rise in ordinal aggregation stability, while still having a significant impact on stability of within group rankings.

Bibliography

Arrow, Kenneth (1953) *Social Choice and Individual Values* (New York: Wiley).

Barabas, Jason (2004) 'How Deliberation Affects Policy Opinions', *American Political Science Review*, 98

Black, Duncan (1948) On the rationale of group decision-making. *The Journal of Political Economy*, 56(1), 23.

Dryzek, John S., and Christian List. "Social choice theory and deliberative democracy: a reconciliation." *British Journal of Political Science* 33.1 (2003): 1-28.

Farrar, C., Fishkin, J. S., Green, D. P., List, C., Luskin, R. C., & Paluck, E. L. (2010). Disaggregating deliberation's effects: An experiment within a deliberative poll. *British journal of political science*, 40(2), 333-347.

Fishkin, James S. (1991) *Democracy and deliberation: New directions for democratic reform* (Vol. 217). New Haven: Yale University Press.

Jacobs, Lawrence R. , Fay Lomax Cook, and Michael X. Delli Carpini. *Talking together: Public deliberation and political participation in America*. University of Chicago Press, 2009.

List, C., Luskin, R. C., Fishkin, J. S., & McLean, I. (2013). Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls. *Journal of Politics*, 75(1).

Luskin, Robert C. & James S. Fishkin (2002, March) Deliberation and 'better citizens' Paper presented at the Joint Sessions of Workshops of the European Consortium for Political Research , Turin , Italy.

David Miller (1992), Deliberative Democracy and Social Choice. *Political Studies*, 40: 54–67. doi: 10.1111/j.1467-9248.1992.tb01812.x

Mojzisch, A., Schulz-Hardt, S., Kerschreiter, R., and Frey, D. (2008) Combined Effects of Knowledge About Others' Opinions and Anticipation of Group Discussion on Confirmatory Information Search. *Small Group Research* April 2008 39: 203-223, doi:10.1177/1046496408315983

Riker, William H. (1982) *Liberalism Against Populism*. San Francisco: W. H. Freeman.

Sunstein, Cass R. (2002), The Law of Group Polarization. *Journal of Political Philosophy*, 10: 175–195. doi: 10.1111/1467-9760.00148