Organizing Deliberation Optimally: A Prelude to the DELMOD Experiment

André Bächtiger
Dominik Wyss

Department of Political Science
University of Lucerne
Frohburgstrasse 3
6000 Lucerne
Switzerland

andrebaechtiger@unilu.ch
dominik.wyss@unilu.ch

Paper prepared for the ECPR General Conference, Bordeaux, 5. -7. September 2013

Very Rough Working Draft – Please do not quote without permission
How is deliberation organized optimally in order to yield its desired effects? As many deliberationists have emphasized, the secret of success may lie in the very process of deliberation and its resulting dynamics. Consequently, we need to think about optimal process designing. Following a recent debate (Manin 2005; Bächtiger 2011; Ani 2013; Bächtiger and Gerber 2014), we propose that properly organized argumentative contestation yields a number of desired products of deliberation, such as epistemic advancement, clarification of positions (Knight and Johnson 2011), intersubjective consistency (Niemeyer 2007), higher levels of integrative complexity (Gruenfeld et al. 1998) and preference changes in directions of workable agreements (Eriksen 2009). Yet, argumentative contestation is not a quasi-automatic feature of a deliberative process. A batch of psychological factors – ranging from satisficing logics to conflict avoidance - may hinder the full deployment of contestation in group discussion. This requires that we deliberately build contestatory practices into deliberative practice. We do this in the form of devil’s advocacy. Using an experimental approach, we test the effects of four communication formats: 1) a sophisticated devil’s advocacy format with a constructive advocate trying to stimulate in-depth reflection; (2) a standard devil’s advocacy format with a purely challenging advocate; (3) an “agent format”; and (4) a free-discussion format without the infiltration of agents. The lab experiment focuses on the Euro crisis and financial bailouts and will be run as an anonymous chat discussion. This paper is a prelude to our DELMOD experiment (which will be run this fall at the University of Vienna in collaboration with Bernhard Kittel and Susumu Shikano): we first present our theoretical rationales why we think that contestation outperforms standard discussion (as we find it in most deliberative citizen events) and then present the experimental setup.

**Communication Modes and Optimal Forms of Deliberation**

When it comes to organizing small group discussions, DPs and many other deliberative events have adopted a format that James Fishkin calls “systematic discussion” (Bernard Manin (2005) also calls it “interactive discussion”). The goal of this format is to attain a systematic and balanced evaluation of the issue under consideration, whereby “competing considerations” – presented to participants in the information material in the form of pros and cons – are put to discursive scrutiny. Facilitation ensures that discussions are on-topic, civil, and not dominated by specific individuals. It is our contention that systematic or interactive discussion is not the most effective communication format, especially when it comes to realize deliberation’s crucial epistemic goals (see Mansbridge et al. 2012).
A thin quality standard that all deliberative democrats embrace is the thorough evaluation of an argument or demand, i.e., an in-depth and unbiased evaluation of its merits and its downsides. A thorough evaluation of an argument or demand also forms the basis of the diverse epistemic and transformative benefits alleged to deliberation. Our claim is that a thorough evaluation of an argument or demand is better realized in a contestatory format than in a discussion format. To be sure, a discussion format is not devoid of contestation, since the very structure of a deliberative process – the succession of arguments and counterarguments – has contestatory elements by default. But the problem here is not one of presence or absence of contestation, but one of basic orientation and quantity: our hunch is that the systematic discussion privileges communication forms other than controversial argumentation, namely the pooling of information, the sharing of experiences, the elaborations of positions which are not sharply polarized or formally adversarial, as well as the identification of converging arguments and common ground. In other words, the discussion format may under-produce conflicting arguments.

To understand why contestation is expected to outperform discussion in this regard, we need to take a look at the (psychological) mechanisms inherent in different communication modes. We start with the conventional discussion format (see Manin 2005). First, controversial arguments may well be present at the beginning of a discussion, but these may be supplanted by consensual arguments that come up in the ensuing discussion. The key problem here is a “satisficing logic”. Since information search is always costly, discourse participants may stop the search for information once a seemingly good reason has been advanced (Manin 2005: 10). Second, the concept of motivated reasoning (Taber and Lodge 2006) suggests that people are unable to control their preconceptions, even when they try to be objective. Consequently, arguments that others find reasonable must resonate with preexisting beliefs and attitudes of others. This may lead to the quick discrediting of discomforting and unpopular arguments which contradicts one’s own views. Third, group discussion always contains the danger of confirmatory bias in that shared information is more frequently mentioned than unshared information. This will “generate a disproportionate amount of information and arguments reinforcing the already prevailing belief” (Manin 2005: 12). Fourth, the interactive structure of discussion allowing every participant a turn may hinder the full deployment of contestation as well. Empowered participants may want to speak up and contribute to the discussion. The consequence may be that evolving confrontations are interrupted, the focus of the debate shifted and an evolving controversy defused. Fifth, the
interactive structure of discussion may also lead to a bifurcation of discourse (Pritzlaff and Nullmeier 2013): new arguments and new proposals are constantly introduced into discussion, which not only enhance the sheer amount of information, but may also come at the price of an in-depth inquiry of the merits and downsides of the arguments and proposals. Sixth, systematic discussion may also experience consensual drifts. When stakes in discussion are not high, participants generally do not enter group discussion with a competitive mindset, trying to convince others that their ideas are superior and win the debate. This can induce participants to vie out for consensual solutions and agreement which, in turn, may have a reductionist logic to argumentation’s content. From a psychological point of view, a consensual drift in a group has also the effect that participants may not look into potential downsides of other arguments “for fear of being perceived as an opponent of a measure objectively promoting the common goal” (Manin 2005: 10). In all these instances, the systematic discussion format may produce a biased evaluation of the merits and downsides of an argument or position.

Contrast this with the mechanisms inherent in a contestatory format. Contestation has a deepening logic to argumentation’s content. Speakers have every incentive to challenge other participants’ arguments and unravel problematic or diffuse links between premises and conclusions. By forcing others to react to these challenges, contestation will help that all sides of an issue are fully explored, that unshared information is uncovered, and that confirmatory bias is reduced. Finally, learning theory suggests that simply reading or hearing information is less effective than actively using it. As such, only direct confrontations – rather than merely ‘hearing’ counterarguments or reading lists of pro and con arguments - might lead people to take proper notice of the disagreement; and, by forcing them to give answers, people might also begin to seriously reflect about the merits and downsides of counterarguments. In sum, contestation creates incentives to challenge each other’s positions and arguments in a radical fashion, creating a deepening logic to argumentation’s content. It is this deepening logic of competitive argumentation which is conducive to a thorough evaluation of the problem at hand.

There are, however, a number of important downsides of competitive argumentation. A first downside is that disputants may also have incentives to distort the issue by inflating or misrepresenting facts and figures. If this happens, contestation can backfire, and participants merely become more entrenched in their pre-discussion positions. A second downside is that contestation can be unpleasant: not every person reacts to contestation and adversarial
arguments in the same way: some people do not mind of being challenged, while others are put off by it. Educational philosophers (e.g. Ikenuobe 2001) have long emphasized that pure adversarialism might be counterproductive by making people retrench and dismissive of a challenge. Practitioners of cross-examination, too, argue that “bullying” and a “harsh style” may be detrimental to the advocate’s success, causing her valid points to be won on the jury (see Brown 1987). Finally, a debate format may also increase the inequality of disadvantaged groups. Psychologists have shown that there is gender bias when roles such as the devil’s advocate are assigned. Women’s reputations can be harmed when they challenge men (Sinclair and Kunda 2000). Similar results were found when Kunda et al. (2002) examined reactions of whites to statements by blacks.

Thus, in order to make contestation effective, it needs to be properly organized. First, we need to ‘institutionalize’ conflict. As Herbert and Estes (1977) note, the institutionalization of conflict and dissent “may help de-personalize the conflict generated by criticism.” (p. 665). To do so, we build a devil’s advocate into the decision-making process, who has an institutional role to argue for an unpopular position (Herbert and Estes 1977). However, the setup of our devil’s advocate is distinctive: it primes on questioning rather than simply arguing for another position and thus accords with the Socrates elenchus. The Socratic elenchus is about “testing or examining the knowledge or wisdom of those reputed (by themselves or others) to be wise” and “showing those who are not wise their ignorance”. Socrates gives an example at his trial when he cross-examines Meletus, one of his accusers. Meletus states a thesis, as something he knows to be true because he is wise about the matter in question. Socrates then asks questions, eliciting clarifications, and extensions of the thesis. He then claims that the original thesis is logically inconsistent with something affirmed in these further responses. For Socrates, it follows that the respondent did not know what he was talking about: true knowledge would prevent one from such self-contradiction. Our devil’s advocate follows this Socratic logic and continuously asks challenging questions to other panel members. While the devil’s advocacy format does not fall prey to the problem of distorting the issue by inflating or misrepresenting facts and figures - our devil’s advocate only asks challenging questions -, this may still not solve the problem that panel members may feel offended and become even more entrenched in their own positions. We try to overcome this tendency by two devices. One the one hand, our devil’s advocate will slowly build-up a fully-fledged examination of issue at hand. In this regard, Brown (1987: 100) mentions an example from the practice of cross-examination: “A masterful cross-examiner,
Kiendl would invariably treat an adverse witness with the greatest courtesy and friendliness, making the witness less suspicious and more relaxed. Slowly, definitely, methodologically, question after question, until the witness was impaled in the meat hooked by his own answers.” On the other hand, we will compare two sorts of devil’s advocates: one acting in a purely challenging way, and one acting in a more constructive way. The latter’s goal is to build bridges to the existing positions and opinions of panel members, inducing reflection and stimulating a productive controversy.

To date, these questions have been rarely put to systematic empirical scrutiny. Nonetheless, existing empirical research provides some evidence that contestation may have deliberative virtues. In a laboratory study, Schweiger et al. (1986) compared the effectiveness of the dialectical inquiry, devil’s advocacy, and consensus approaches to strategic decision making by groups. Results showed that both dialectical inquiry and devil’s advocacy led to a higher level of critical evaluation of assumptions and better quality recommendations than the consensus treatment. This result is even more remarkable since in the consensus treatment participants were asked to consider counterarguments carefully and critically as well as to avoid changing their mind simply to avoid conflict and reach agreement. Similar results are found in citizen juries. In juries oriented towards consensus, Huitema et al. (2007) found tendencies that not every argument was properly discussed: “We observed that this led the juries to avoid both potentially divisive topics and, to some degree, the adoption of unreal assumptions in their recommendations.” (p. 302) In a recent evaluation of communication modes in a pan-European deliberative poll (Europolis), Bächtiger and Gerber (2014) find that small group discussions in Europolis are mainly instances of “gentlemanly conversation”. Contestation was not absent, but it is neither the most important communication mode nor is it deployed in a bold manner. When a “clash of conflicting arguments” occurred, challengers were rarely insisting but instead tried to charter consensual ground. Evolving controversies were also hindered by the fact that other participants frequently act as “mediators”, shifting the controversy in other directions and thus defused it at the same time. And the interactive discussion was also conducive to a bifurcation of discourse with a relatively shallow scrutiny of arguments and proposals. Focusing on the Australian Citizen Parliament, Curato et al. (2013) observed that a technique imported from mediation – “appreciate inquiry” – led to a high satisfaction of participants with the deliberative procedure, but undermined the critical reflection of the issues under discussion. However, the Schweiger et al. (1986) experiment also showed that subjects in the consensus groups expressed greater acceptance of their
groups’ decisions as well as a desire to continue to work with their groups compared to participants in dialectical inquiry or devil’s advocacy groups. This clearly speaks for a differentiated approach to devil’s advocacy, as suggested above. In sum, while existing empirical research provides hints that contestatory formats may be superior compared to standard discussion formats, the evidence is far from being conclusive: neither are the various theoretical claims tested in a comprehensive manner, nor are obvious downsides of contestation addressed in a productive way. Our experiment sets out to change this.

Predictions

What outcomes follow when participants engage in a process of argumentative contestation, induced by a Socratic-style devil’s advocate? We make six predictions.

Prediction 1 (Epistemic advancement): Devil’s advocate formats, by fully exploring all sides of an issue, for uncovering unshared information, and by focusing on the downsides of specific proposals and arguments, produce a higher level of (in-depth) knowledge than a discussion format.

Prediction 2 (Clarification): A standard devil’s advocate format has a clarification function (Knight and Johnson 2011): after the confrontation, participants know better what they want and consequently better align their basic preferences to concrete policy proposals (“intersubjective consistency”; Niemeyer 2007).

Prediction 3 (Integrative Complexity): A sophisticated devil’s advocate format may also lead to a higher level of integrative complexity compared to a standard devil’s advocate format as well as a discussion format. We expect that the revelation of inconsistencies in one’s position in combination with a constructive attitude on part of the Devil’s advocate is conducive to the acknowledgement and appreciation of other positions and arguments.

Prediction 4 (Opinion Change and Integrative Solutions): Regarding opinion changes, two scenarios might arise in the devil’s advocacy format, depending on the ‘nature’ of the devil’s advocate: (1) if the devil’s advocate is constructive and produces higher levels of integrative complexity (prediction 3), participants may vie out for middle positions, integrative solutions (a sort of “working agreement” (Eriksen 2009) or “linear averaging” (see Hartmann et al. 2008)) (2) if the devil’s advocacy is merely challenging and fails to create a productive controversy (prediction 2), participants may only clarify positions but not change minds and not vie out for integrative solutions. In the discussion format, the outcomes are somewhat open-ended,
too: one possibility is that conformity pressures lead panel members to shift their opinions to the majority opinion in the group; another possibility is that groups experience a consensual drift which promotes integrative solutions. Compared to devil’s advocacy, however, we expect this to be easy integration, with less epistemic gains and a lower level of integrative complexity.

**Prediction 5 (Decision Acceptance):** Following Schweiger et al. (1986), devil’s advocate formats may reduce satisfaction both with process and outcomes, albeit a sophisticated devil’s advocacy format with a constructive advocate might temper this effect.

**Prediction 6 (Inequality):** Devil’s advocate formats are cognitively demanding format, thus enhancing communicative disadvantages of women and other minority groups.

**Experimental Design**

Our DELMOD experiment focuses on the Eurozone bailouts, with the question whether (student) participants are willing to assist crisis states (such as Greece) or not. Compared to existing research (Schweiger et al. 1986), we do not focus on virtual decision cases but on real preferences of (student) participants. The participants are recruited from a large student pool at the University of Vienna, enabling us to maximize preference heterogeneity in the discussion groups. The participants are compensated with a fixed lump sum. Hence, there is little incentive for strategic behavior on part of the participants; otherwise it would be exceedingly difficult to causally relate choices to different communication treatments. The experiment will be run as an anonymous chat discussion in a computer-lab provided by the experiment software z-Tree by Fischbacher (2007). The anonymous and text-based features of the discussions are intended to limit patterns of social influence, a barely recognized (and understudied) problem of traditional face-to-face encounters (Price 2009: 43). By the same token, chat conversations may not be optimal for an exchange of in-depth and complex arguments. Experience with online chat reveals that argumentative complexity tends to be low, especially when compared to face-to-face discussion (Pedrini 2013). In order to stimulate higher quality discussions, participants are explicitly asked to provide in-depth justifications for their positions and to take time to do so.

We use survey techniques to measure the individual preferences with regard to European bailouts, knowledge gains, alignment between basic preferences and policy proposals (intersubjective consistency), integrative complexity, opinion change, and the acceptance of
process and decisions. With regard to measuring participants’ preferences, we apply a conjoint method where participants have to rank five predefined policy packages (see Table 1). The method, while rarely used in political science, has a number of advantages compared to standard preference scales (Bechtel et al. 2013): First, by having to choose among different policy packages, participants are confronted with real-world trade-offs. Second, the conjoint method does not only give answers to one-dimensional questions - whether to support one specific policy or not – but also allows figuring out under what conditions participants are ready to support a specific policy.

Table 1: Conjoint Survey: Packages

<table>
<thead>
<tr>
<th>Packages</th>
<th>Distribution of Costs between Actors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Greece)</td>
</tr>
<tr>
<td>I</td>
<td>75%</td>
</tr>
<tr>
<td>II</td>
<td>40%</td>
</tr>
<tr>
<td>III</td>
<td>30%</td>
</tr>
<tr>
<td>IV</td>
<td>5%</td>
</tr>
<tr>
<td>V</td>
<td>0%</td>
</tr>
</tbody>
</table>

The participants are asked to rank 5 packages. The packages differ in the way how the costs of financial crisis are distributed among the three actors Greece, European Union and private investors. Package I assigns most of the costs to Greece, while in package V all the costs are borne by the European Union.

Experimental Procedure

At the beginning, participants fill out a survey which checks for their initial preferences regarding Eurozone bailouts, tests their knowledge on the issue, and asks them a batch of general questions on justice (see figure 1). This first survey provides the basis for group assignment. All discussion groups consist of five participants, but are slightly biased towards a solidaristic or the non-solidaristic position. To maximize preference heterogeneity in the groups, a matching algorithm is used for assigning participants to groups. After filling out the first survey, participants receive information material on the Euro bailouts, comprising pro and con arguments for a solidaristic position (unconditional help for bailout countries) and for a non-solidaristic position (self-help of crisis states).
Next, four communication formats are implemented: 1) a sophisticated devil’s advocacy format with a constructive advocate trying to stimulate reflection; (2) a standard devil’s advocacy format with a purely challenging advocate; (3) an “agent format”; and (4) a free-discussion format without the infiltration of agents.

Ad (1) and (2): Both devil’s advocacy formats implement a confrontation session, as suggested by Herbert and Estes (1977). The difference between the sophisticated and the standard devil’s advocacy format is that the devil’s advocate role is more constructive in the former than in the latter. The setup is as follows: After ten minutes of open-discussion, the devil’s advocate intervenes in the discussion. The devil’s advocate picks participants with majority opinions and then asks them a set of challenging questions. During the confrontation session, the exchange is restricted to the Devil’s advocate and specific participants. In the standard devil’s advocacy format, the level of contestation increases from each question to the other. This is how intervention looks like:

**Example of an intervention by a standard devil’s advocate:**

<table>
<thead>
<tr>
<th>First Statement: User 3: You argue that Greece alone has to bear the consequences for their own actions. Is this correct, User 3?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second Statement: To put it in other words: You refuse assistance to the Greek people even though these decisions have been made by past decision-makers?</td>
</tr>
<tr>
<td>Third Statement: Why do you refuse assistance to people in need (like young, poor, and unhealthy people), when the crisis can be blamed to former decision-makers? Why do you want to punish innocent people?</td>
</tr>
</tbody>
</table>

In the sophisticated devil’s advocacy format, contestation is complemented with a constructive component, whereby the devil’s advocate, after having asked a batch of challenging questions, tries to build bridges from the existing positions of participants to conflicting arguments. The goal is not mediation (which would be at odds with the concept of devil’s advocacy), but stimulating in-depth reflection among participants.

**Example of an intervention by a constructive devil’s advocate:**

<table>
<thead>
<tr>
<th>First Statement: I do not like the fact that the Greek population is bearing the consequences of the crisis, which is caused by some former decision-makers. How can we avoid such unfair situation?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second Statement: I see your point against any financial assistance. However, in case of state bankruptcy, we have to think about how to support people in need.</td>
</tr>
<tr>
<td>Third Statement: I feel obliged to assist at least the poorest. Are there possibilities to ensure that elementary institutions like schools and hospitals will not be seriously affected by the crisis?</td>
</tr>
</tbody>
</table>
After the confrontation session, participants have ten additional minutes to continue discussion. The expectation is that participants will pick up the points made in the confrontation session, leading to a highly focused, in-depth and potentially creative discussion. In particular, we expect participants with minority views to feel empowered to engage in the debate as well.

Ad (3): In order to isolate the effect of the devil’s advocate and control for the presence of an additional person in the group, we need to implement another communication format, dubbed “agent format”. The setup of the agent format is very similar to the devil’s advocacy formats: it involves 15 minutes of open-discussion before an agent intervenes on behalf of the minority. The main difference is how information is presented: The systematic agent does neither single out a participant nor asks questions directly. The agent only reinforces the arguments of the minority by stating them. Here is an example of an agent intervention:

<table>
<thead>
<tr>
<th>Example of an intervention by an agent:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Statement:</strong> The young, poor and unhealthy are the ones who suffer most from insolvency.</td>
</tr>
<tr>
<td><strong>Second Statement:</strong> I think that the responsibility for the debt crisis rests with some former decision-makers and not with the Greek people.</td>
</tr>
<tr>
<td><strong>Third Statement:</strong> To me, refusing any assistance to people in need is not a valid option.</td>
</tr>
</tbody>
</table>

To exclude informational effects, it is important that the statements of the agent cover the same information as the statements of the devil’s advocate formats. In sum, since agent interventions are likely to influence the discussion in some ways (e.g., by highlighting some specific arguments, coming up with new information or simply manipulating the dynamics of the talk), the agent format helps us to control for such effects.

Ad (4): the “free discussion format” is not infiltrated by any agents and serves as a baseline treatment. It partly conforms to the usual free discussion setup in economic experiments (see Goeree and Yarif 2011). The difference is that we are interested in the effects of argumentation, not only in the coordination and information function of talk. Consequently, participants are explicitly asked to justify their positions. In combination with the information material, the free discussion format also mimics Fishkin’s systematic discussion. The four communication formats are summarized in table 2.
Table 2: The four communication formats

<table>
<thead>
<tr>
<th>Constructive Devil’s Advocacy format</th>
<th>Standard Devil’s Advocacy format</th>
<th>Agent format</th>
<th>Free discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Devil’s Advocate asking questions to participants of the majority, but with a constructive component.</td>
<td>In contrast to the former, this Devil’s Advocate asking purely challenging questions.</td>
<td>The agent reinforces the minority by presenting minority’s arguments.</td>
<td>The systematic discussion groups are solely composed of participants.</td>
</tr>
</tbody>
</table>

The experiment is split into two rounds (see figure 1): in the first round participants only discuss the merits and downsides of Eurozone bailouts. Our four treatments are implemented in this round. At the end of first discussion round participants are asked to indicate their individual conjoint-decision. In the second phase, participants work out a collective position, whereby they are explicitly asked to strive for a common position if possible (albeit there is no monetary incentive to achieve a common solution). In this round, no intervention by an agent will occur. The preparation of a collective position lasts 10 minutes (maximum), but if participants have found agreement on a position, they can finish early. We expect that groups exposed to devil’s advocates might take longer to come to a decision than groups in the agent or free discussion formats. At the end of this round each participant is asked to enter the result of the collective decision making process. To check whether the group has identified a common position, all group members must have answered in identical ways.
The experiment ends with a final questionnaire. Here, we re-check for knowledge on the issue but also check for integrative complexity. With regard to integrative complexity, we focus on the reasons underlying participants’ choices. We ask them to rate the pro and con arguments for a solidaristic position (unconditional help for bailout countries) and for a non-solidaristic position (self-help of crisis states).

References


