



Disagreement and Deliberation: Evidence from Three Deliberative Mini-Publics

Kimmo Grönlund¹ · Kaisa Herne² · Kim Strandberg³ · Peter Söderlund¹

Accepted: 30 June 2021
© The Author(s) 2021

Abstract

This article is based on three experiments in citizen deliberation. We ask whether disagreement at group level as well as at individual level influence participants' experiences of deliberation. In all three experiments, participants discussed in small groups and answered surveys before and after deliberations. The experiments were population-based with random selection. The topic of the first deliberation was nuclear power, the second dealt with immigration, and the third concerned policies for a language spoken by a national minority. The degree of group level disagreement was subject to experimental manipulation. In the first experiment, all the participants discussed in groups with mixed opinions. In the second experiment, participants were first categorized according to their baseline views, and then randomly allocated into either mixed or like-minded groups. In the third experiment, everyone discussed in like-minded groups. A trained facilitator moderated all small group discussions in the first two experiments. In the language experiment, the participants were randomly assigned into two treatments: groups with both moderation and deliberative norms, and 'placebo' groups. Our dependent variables consist of participants' self-reported experiences of being heard in the discussion, and their feelings of mutual respect. The results show that all participants—regardless of group level disagreement—tend to be satisfied with deliberation. The only exception is the first experiment, where disagreement decreased process satisfaction slightly. At the individual level, participants' deviation from the group mean had almost no effect.

Keywords Deliberative democracy · Citizen deliberation · Experiment · Disagreement · Satisfaction

✉ Kimmo Grönlund
kimmo.gronlund@abo.fi

¹ Social Science Research Institute and Political Science, Åbo Akademi University, ASA-building, Fänriksgatan 3, 20500 Åbo, Finland

² Faculty of Management and Business, Politics, Tampere University, Tampere, Finland

³ Social Science Research Institute and Political Science, Åbo Akademi University, Vasa, Finland

Introduction

The normative model of deliberative democracy holds that discussion in heterogeneous groups, respect for others, reason-giving and reflection are essential parts of legitimate political decision-making (Dryzek, 2000; Elster, 1998). In this article, we focus on the first part, i.e. heterogeneous groups, and ask about opinion diversity, particularly in how it influences participants' experiences of deliberative discussions. While diversity is often perceived as an essential element of deliberation, there are also those who emphasize deliberation among like-minded or homogenous groups as a mechanism for the empowerment of marginalized groups (Karpowitz et al., 2009). In the present article, deliberation is defined as structured discussion guided by a moderator and specific discussion rules. Like-minded deliberation may be a part of a deliberative system where a diversity of perspectives is achieved at the system level but not in each individual discussion (Mansbridge et al., 2012). However, we have little knowledge of how opinion diversity, or the lack of diversity, influences participants' experiences, which in turn may be highly relevant when it comes to their willingness to engage in subsequent political discussions in general and deliberation in particular. So far, the main empirical evidence on how disagreement within a group affects participants in structured deliberation comes from Esterling et al. (2015). Their findings suggest that the level of disagreement in deliberation should neither be too low, nor too high. We want to assess if this holds in other contexts and ask what consequences the degree of disagreement has on those who deliberate. By analyzing the impact of disagreement at both the group and individual level, we test if people consider participation as a pleasant experience and whether they would be ready to participate in a similar event anew in the future.

A study of the consequences of the level of disagreement in a deliberative setting is important because sometimes like-minded deliberation may be justified. While the deliberative ideal emphasizes a diversity of opinions, deliberation among disempowered groups may enable group members' preference formation and enhance their mobilization. To evaluate the potential of like-minded deliberation to achieve these ends, it is important to know how participants experience deliberation in homogenous groups. If discussion is not perceived as a pleasant experience, it is unlikely that people would be willing to take part again. This would undermine enclave deliberation as an effective tool of mobilization and inclusion of marginalized groups.

We use a unique set of data from three separate experiments (total $n=542$) designed to test the consequences of deliberation. The experiments share most features, in particular the focus on small group deliberation, but they concern different topics, nuclear power, immigration and the status of the Swedish language in Finland. The data enable us to compare the consequences of deliberation in different settings. The degree of disagreement is measured in relation to the small groups where the participants discussed. We look both at the overall disagreement within the group, and individual participants' deviation from the group mean. Overall, our results suggest that deliberation is a pleasant experience for participants, and that disagreement does not have an impact. Only in the first experiment high levels of

disagreement lead to less satisfaction with the process, both at group and individual levels.

Previous Research and Research Question

The normative idea of deliberative democracy requires participation or representation of all those who are affected by a collective decision, which usually entails a diversity of opinions (Gutmann & Thompson, 1996, p. 128). Thus, the legitimacy of democratic decision-making requires that all affected interests and perspectives are fairly considered in the deliberative process. In addition, exposure to cross-cutting perspectives has certain instrumental or, more precisely, epistemic benefits. Diversity in deliberating groups encourages people to correct their own biases of reasoning and enhance their capacity to consider a variety of perspectives (Mercier & Landmore, 2012; Morrell, 2010). Empirical evidence also demonstrates that deliberation among people with conflicting opinions enhances their level of political knowledge and ability to see other perspectives (e.g. Andersen & Hansen, 2007; Grönlund et al., 2017).

In contrast, negative aspects have been associated with deliberation occurring in a homogenous group. Karpowitz et al. (2009) put forward the main arguments related to the problems of like-minded deliberation. They start with the argument that homogeneity undermines democratic legitimacy because it leads to a failure to consider the common good and all affected interests in the deliberative process. According to Karpowitz et al. (2009), a homogenous group is also likely to limit the diversity of different perspectives in deliberation, which undermines the epistemic benefits of deliberation based on the correction of biases and mutual learning. For example, Sunstein (2002) claims that discussion in homogenous groups, or opinion enclaves, will lead to a polarization of opinions, a limited information pool and even an amplification of cognitive errors. Group polarization refers to a process where a like-minded group becomes more extreme because the arguments supporting the dominant position are reinforced rather than challenged in the discussion.

It is noteworthy that the heterogeneity or homogeneity of a group can be defined in several ways. In addition to opinions or attitudes, it can reflect the demographic background of group members, or identity. Karpowitz et al. (2009) point out that deliberation in a homogenous group can benefit certain disempowered groups (see also Abdullah et al., 2016; Himmelroos et al., 2017). For these types of groups, it might be easier to articulate their specific needs or interests in like-minded groups. Further, deliberation in a like-minded group can be an efficient way for a disempowered group to get its voice heard in a wider public debate, and in this way enclave deliberation can promote more inclusive decision-making. This is particularly possible if enclave deliberation is connected to deliberation among a wider public with a diversity of views. In line with this, Mutz (2002) shows that discussion among like-minded people increases the willingness for political participation, whereas cross-cutting communication may undermine it. However, the evidence on the influence of opinion diversity on political mobilization is mixed and suggests that the

relationship can be conditional on various contextual factors (Kwak et al., 2005; McClurg, 2006; Pattie et al., 2009).

Experimental evidence supports the view that deliberation in enclaves does not necessarily lead to the negative consequences pointed out by Sunstein (2002). Karpowitz et al. (2009), for instance, show that taking part in a consensus conference increased knowledge over the discussed topic, self-efficacy and interpersonal trust, and did not lead to a polarization of opinions among certain marginalized groups. In their study, enclaves were based on a set of background variables, such as income or ethnic background. Grönlund et al. (2015), using the same immigration experiment data as we use here, manipulated the group composition of a deliberative discussion according to pre deliberation opinions. Some of the participants deliberated in like-minded groups and others in mixed groups with heterogeneous opinions. The main observation was that opinions did not polarize even in the like-minded groups and that knowledge over the deliberated issue increased in both types of groups. Moreover, increased out-group empathy was observed to some extent even in like-minded groups (Grönlund et al., 2017).

What both Karpowitz et al. (2009) and Grönlund et al. (2015) have in common is that they study discussion that took place under the structured conditions of deliberation. Karpowitz et al. (2009) report the results of taking part in a consensus conference, one of the commonly used forms used to organize deliberative discussions. Further, in the Grönlund et al. (2015) experiment, all discussions took place in conditions that are supposed to enhance the deliberativeness of the discussion. Indeed, Grönlund et al. (2015) suggest that it is these types of deliberative norms that can alleviate group polarization tendencies. A subsequent experiment (Strandberg et al., 2019) finds evidence in support of this interpretation. In the experiment, like-minded small groups with facilitators and discussion rules were compared with groups without facilitation and rules. The results show that polarization occurred in the latter groups but not in the former.

While there is some evidence on the influence of discussion on heterogeneous versus homogenous environments on people's opinions, the topic is nevertheless understudied. Further, the existing literature mainly focuses on opinion and knowledge changes, perspective taking and civic virtues, such as trust and efficacy. We pay attention to the consequences of group level disagreement on the experiences of participants taking part in a deliberative discussion. While a certain level of disagreement is deemed desirable for a multifaceted deliberation, people may in fact find it unpleasant to discuss issues in environments with high levels of disagreement (Theiss-Morse & Hibbing, 2005). Esterling et al. (2015) thus suggest that a *medium level of disagreement* is ideal for good deliberation. They argue that the deliberative ideal in fact requires some disagreement: "with no disagreement, reasons need not be offered nor considered, and with too much disagreement reasons fall on deaf ears" (Esterling et al., 2015, p. 530). They test whether a moderate level of disagreement enhances satisfaction with deliberative discussion. They study survey responses from *California Speaks* deliberations held on the healthcare reform in 2007. In the event, participants engaged in structured deliberation with trained moderators. Certain rules of discussion declaring the characteristics of ideal deliberation, such as listening respectfully to others and not dominating the discussion, were

followed in the process. Esterling et al. (2015) look at the influence of table level disagreement, i.e. disagreement in the small-n group, on participants' satisfaction with the deliberative process. Their main observation is that participants at those tables where disagreement was moderate indeed had the highest levels of process satisfaction. Based on this result, deliberation appears to work best when the group level disagreement is neither too low nor too high.

We test the suggestion of Esterling et al. (2015), albeit with a slightly different design. We look at evidence from three experiments designed to engage subjects in deliberation. The discussions took place in small groups and they were moderated and enforced by specific discussion rules. Our study adds to the research of Esterling et al. (2015) by two features. First, we are able to get data from deliberations where the level of disagreement within small groups was manipulated in a controlled manner. Further, we are able to look at three different topics of deliberation and study whether the type of topic might influence the subjects' experiences of the process. Finally, with data from the third experiment, we are also able to test the impact of facilitation and rules since that was varied so that half of the groups lacked facilitation and rules, whereas the other half had these in place. We ask whether the observation of Esterling et al. (2015) also holds in the three different deliberative settings we study. The advantage of our research design is that we may compare three different contexts because the experience of a deliberative discussion may depend, in addition to opinion diversity, on the specific small group dynamics, the topic of deliberation, payment for participation and some other contextual factors. Since a lack of theorizing and a limited amount of previous evidence does not provide a basis for hypothesis testing, we therefore pose the following research question: *How much disagreement is good for deliberation, when measured through participants' experiences?*

Experimental Procedures

We focus on three separate experiments that share certain characteristics but also differ in some respects to further test the influence of disagreement on satisfaction under varying conditions. All three experiments were designed to examine the consequences of taking part in a controlled deliberative discussion. The first experiment, consisting of 12 small groups (10–13 participants per group, total $n=135$) concerned nuclear power and energy policies. The second experiment concerned immigration and had 26 small discussion groups (6–9 participants per group, total $n=207$). The third experiment had 31 small discussion groups (4–9 participants per group, total $n=202$) and the topic for deliberation was the Swedish language in Finland.¹ In the first two experiments, participants were based on a regional random sample, whereas participants in the third experiment were recruited through a random sample of the whole Finnish population. In each case, the experiences of

¹ Finland is a bilingual country and Swedish is an official language. It is spoken by 5.2 percent of the population.

having taken part in the deliberative event were measured through a survey right after the discussion. We use the questions pertaining to the experiences of taking part in the event as dependent variables.

The experiments were originally designed to test the influence of a certain manipulation in the conditions of deliberation. The first of these conditions dealt with the composition of the discussion group. In the first experiment, all discussion groups were mixed in the sense that the participants' opinions on nuclear power and energy policies varied within each group. In the second experiment on immigration, the composition of the discussion group was manipulated so that a part of the small groups was mixed in terms of the participants' opinions, whereas the remaining groups consisted of like-minded participants. We created the enclaves with the help of a pre-test survey, measuring the respondents' opinions on immigration. Respondents with negative attitudes to immigration formed a *con* enclave, and respondents with a positive view on immigration formed a *pro* enclave. Within these enclaves, subjects were randomly assigned either into mixed or like-minded groups for deliberation. In the third experiment on the Swedish language, opinion enclaves were also formed on the basis of the first survey, but all discussion groups were within like-minded opinion enclaves. Thus, recruitment in the second and third experiment was different from the first experiment where pure random sampling was used. In the latter two, respondents whose opinions were not clearly in favor or against immigration (2nd experiment) or the Swedish language (3rd experiment) were not included in any of the enclaves. These undecided persons, whose baseline views were close to the mean on the index variable measuring opinions on the issue at stake were not invited to deliberate.

The second condition that varied between the first and the other two experiments was how post deliberation opinions were measured. In all cases, the participants were surveyed on their opinions regarding the issue at hand directly after deliberation. In the first experiment, however, all small groups made a decision or wrote a statement. Half of the groups made a decision with a secret ballot, whereas the other half wrote a common statement on whether to build more nuclear power. The common statement groups were asked to formulate a statement which all group members could accept. However, it was also emphasized that consensus should not be forced and if consensus was not reached, the statement should simply indicate the number of individuals for or against a certain view. In the second and third experiment no statements or decisions were made at group level.

The third condition that we varied across the experiments was the use of facilitation and discussion rules in the discussion groups. In the first two experiments, this condition was constant since all groups were subject to facilitation and discussion rules. The third experiment, however, varied the use of facilitation and rules so that half of the groups used it and the other half discussed without rules and facilitation. The aim was to examine further whether a deliberative treatment, discussion with facilitation and rules, differs from a non-deliberative treatment, free discussion without facilitation or rules. In the deliberative treatment, a trained facilitator guided the discussions and implemented discussion rules derived from deliberative norms, whereas in the free-discussion treatment, a member of the research team was present in the small group discussion albeit he or she remained passive and did not facilitate

the discussion. The discussion rules supported the norms of reasoned justifications, reflection, sincerity and respect, whereas the facilitation is supposed to enhance reciprocity, inclusion and equality of discussion. Table 1 presents the main characteristics of the three experiments.

Table 1 shows that apart from the experimental treatment in each experiment (decision-making method, group composition, deliberative norms), all other factors were held constant throughout.² Looking at the more detailed procedures of the deliberations, there was only one additional substantial difference between the experiments, namely the fact that the nuclear power experiment had a hearing session of an expert panel. The panel consisted of four persons; two men and two women, two MPs, a lobbyist for nuclear power companies and a representative for an environmental NGO. The participants heard and questioned the expert panel in a plenum after having read the information package but before going to their small groups for deliberation. In the experiment on immigration, the information package was presented in a plenary session, but there was no expert panel. In the deliberation about the Swedish language, an information package was handed out in the discussion groups and no expert panel was heard.

Otherwise, the experiments followed comparable procedures. Participants were first contacted with a pre-deliberation survey and an invitation to take part in the mini-public. When the participant pool was known, they were mailed a brief information package containing the basic facts on the topic. All the experiments took place in campus buildings. When participants arrived at the venue, they were welcomed, instructed and asked to fill in a knowledge quiz (there was no pre deliberation knowledge quiz in the third experiment). After this initial plenary session, participants gathered in smaller rooms in their own group. Discussions were interrupted with coffee and lunch, and at the end of the day, participants filled in the post deliberation survey. In the nuclear power mini-public, discussions lasted about 3 h, in the immigration experiment, they lasted about 4 h, and in the Swedish language experiments about 2 h. In the nuclear power and immigration experiments, participants filled in a mailed follow-up survey. In the Swedish language and nuclear power experiments, debriefing material was sent to the participants, whereas a debriefing event was organized in the immigration experiment. None of the mini-publics had a direct impact on policy-making processes, and participants were informed that they would take part in a research project. However, the processes were planned to follow the basic tenets of mini-publics, i.e. a random selection of invited participants, balanced information on the topic, and moderated small group discussions primed with specific rules. Even though there was no direct policy-link, all the experiments received some media attention, i.e. the research team produced press releases that included the main results. Table 2 represents the phases of each experiment.³

² All the experiments also exploited quasi control groups, which did not gather or discuss at all. In the first two experiments, the control group filled in pre- and post-surveys, in the third experiment the control group only filled in one survey.

³ For a more detailed description of procedure at each experiment, see Setälä et al. (2010), Grönlund et al. (2015) and Strandberg et al. (2019).

Table 1 Characteristics of three deliberation experiments

Topic	Year	Sample	Participants (n)	Small groups (n)	Participants in small groups (n)	Group composition	Decision making procedure	Facilitation and rules of discussion
Nuclear power	2006	Regional	135	12	10–13	Mixed groups	Vote/common statement	Yes
Immigration	2012	Regional	207	26	6–9	Mixed and like-minded groups	None	Yes
Swedish language	2014	National	202	31	4–9	Like-minded groups	None	Yes/no

Table 2 Phases of the experiments

	Nuclear power	Immigration	Swedish language
Pre deliberation survey(s)	T1	T1+T2	T1+ T2
Deliberation day	November 18, 2006	March 31–April 1, 2012	November 2014
Knowledge quiz	T2	T3	-
Information package	Yes	Yes	Yes
Expert panel	Yes	–	–
A short poll	T3	–	–
Small group discussions	3 h	4 h	2 h
Post deliberation survey	T4	T4	T3
Follow up survey	T5	T5	–

What is relevant from the perspective of this paper is that the experiments vary in terms of the small group composition and the surrounding circumstances. In particular, the level of disagreement within the groups varied because of the design of the experiments. The ideal of deliberative democracy holds that participants should be exposed to a variety of opinions. Thus, in the experiment on nuclear power, all participants discussed in groups with mixed opinions which, a priori, ought to increase the level of disagreement in the group. In the second experiment on immigration, the participants were first assigned into *pro* or *con* enclaves according to their immigration opinions. Thereafter, they were randomly assigned into like-minded or mixed small groups where the discussion took place. The subjects were not informed about the manipulation of the group composition. The second experiment thus contained a mix of conditions inducing disagreement or not. A similar procedure was used in the third experiment where two opinion enclaves were first formed, and the participants were thereafter randomly allocated to treatments within the enclaves. The third experiment thus does not contain any groups with mixed opinions.

It is noteworthy that in addition to our experimental treatments, the topic of deliberation varied between the three experiments. Different topics may influence participants' experiences of the process.⁴ An important difference is between *cold* deliberation which takes place when stakes are low, and *hot* deliberation where stakes are high (Fung, 2007). It is possible that topics, which participants consider "hot" have higher influence on process satisfaction compared to "cold" topics. Ideally, we should therefore have designed an experiment where different procedural or group composition factors are varied, whereas the topic is held constant, or alternatively, held everything else constant and vary the topic. Since the experiments were not originally designed to study only the questions focused on here, we do not have this option. However, with the data we have, we can compare treatment groups *within*

⁴ We wish to thank one of the anonymous referees for pointing this out.

each experiment, and with the selection of three experiments we are able to get data from *different* deliberative contexts.

Measures of Disagreement and Process Satisfaction

In order to establish the level of disagreement at both the group and individual levels, we exploit a number of statements presented primarily on Likert scales relating to the topic of each experiment. Indices are first created by calculating the arithmetic means of the participants' opinions on the discussed topic in the three experiments. In the case of nuclear power, we use eight items, which load strongly on a single factor. Principal component factor analysis was conducted using a polychoric correlation matrix. In a similar manner, for both the experiments on immigration and the Swedish language we use 14 items that measure the participants' attitudes on each topic before actual deliberation. For reasons of comparability, the responses were recoded to range between 0 and 10. Missing data on one variable are replaced with that variable's grand mean of all participants. This is acceptable since data appear to be missing completely at random. All analyses were performed with Stata 16.

In line with Esterling et al. (2015), two measures of disagreement are created. *Group level disagreement* is operationalized as the standard deviation of the arithmetic mean of the index variable which measures the participants' pre-deliberation opinions. This is calculated for each small group separately. Since person i assesses the group context, person i 's own position is not included when group level disagreement is calculated (see Esterling et al., 2015, p. 537). The higher the standard deviation, the more the participants' opinions deviated from each other before deliberation. Table 3 displays the group level disagreement scores in all three experiments. It also shows the group means for each group on the classification variable, i.e. the index variables for opinions on nuclear power, immigration and the Swedish language. *Individual level disagreement* is measured as an individual participant's distance from the group mean for the index variable. We use the absolute value (i.e. if the value is negative, it is turned positive) so that zero denotes no disagreement and a high value indicates disagreement with the average group member.

Having classified the independent variables, i.e. group level and individual level disagreement, we move on to the dependent variables, i.e. how participants evaluated the deliberative events. We make use of questions that were posed directly after deliberation, when participants still sat at tables in their rooms.⁵ Thus, the timing and setup of the measurement is identical in all three experiments. For the two first experiments, the questions posed are similar and therefore comparable (see Table 4).⁶ The first one is a direct question on whether the participants think

⁵ We replaced the missing values for each satisfaction item with the mean value of the available data. As a robustness check, we later ran regression models where the respondents with one or more missing values for any of the survey items that were used to create the index variables were dropped. The coefficient estimates did not substantially change and therefore we did not choose to drop observations with any missing values.

⁶ These items also resemble those that were included to measure "satisfaction with the quality of the discussion itself", or "process satisfaction", by Esterling et al., (2015, p. 535).

Table 3 Group level disagreement, and group means, in three experiments

Group	Experiment					
	Nuclear power		Immigration		Swedish language	
	Disagr	Mean	Disagr	Mean	Disagr	Mean
1	2.0	4.9	0.5	4.3	0.4	7.7
2	2.1	4.5	0.6	4.1	0.5	7.6
3	2.4	4.6	0.8	3.9	0.6	7.4
4	2.5	2.9	0.8	7.1	0.7	7.3
5	2.6	3.9	0.8	7.4	0.7	6.7
6	2.6	3.9	0.9	8.0	0.7	6.7
7	2.7	3.6	0.9	7.5	0.7	3.6
8	2.8	4.4	1.0	6.9	0.8	8.0
9	2.9	5.3	1.0	7.7	0.8	6.9
10	3.0	4.6	1.0	7.6	0.8	7.3
11	3.3	5.1	1.1	7.7	0.8	7.5
12	4.3	5.2	1.1	7.3	0.8	7.3
13	–	–	1.2	4.0	0.8	7.7
14	–	–	1.3	7.5	0.9	3.6
15	–	–	1.6	5.2	0.9	3.4
16	–	–	1.7	2.6	0.9	8.0
17	–	–	1.8	5.3	1.0	6.8
18	–	–	1.8	4.9	1.1	8.1
19	–	–	2.0	5.4	1.1	3.3
20	–	–	2.1	6.0	1.1	3.6
21	–	–	2.2	5.2	1.1	2.8
22	–	–	2.4	5.3	1.1	3.1
23	–	–	2.4	5.6	1.1	8.0
24	–	–	2.5	4.8	1.2	2.6
25	–	–	2.5	4.6	1.3	7.4
26	–	–	3.1	5.0	1.3	3.1
27	–	–	–	–	1.3	2.8
28	–	–	–	–	1.5	7.3
29	–	–	–	–	1.5	8.0
30	–	–	–	–	1.5	7.4
31	–	–	–	–	2.9	3.6

We measured group level disagreement for each individual i by excluding person i 's value from the standard deviation. Therefore the group level disagreement score may vary between individuals within the same group. In this table, for the sake of illustration, each group level disagreement score is the group average of the standard deviations for all individuals in the group

Table 4 The dependent variable of the study: process satisfaction index, scale from 0 (not satisfied at all) to 10 (totally satisfied)

Experiment and survey items
Nuclear power (Mean = 8.35, SD = 1.30, Alpha = 0.60)
1. Taking part in the citizen deliberation was a pleasant experience. (Mean = 9.36)
2. My knowledge about energy issues increased during the discussion. (Mean = 8.34)
3. My readiness to participate in political and civic activities increased during the discussion. (Mean = 6.26)
4. I would be happy to participate again in a similar event. (Mean = 8.85)
5. I was able to present my views in a satisfying manner during the discussion. (Mean = 8.94)
Immigration (Mean = 7.45, SD = 1.27, Alpha = 0.61)
1. Discussing in a group was a pleasant experience. (Mean = 8.37)
2. My knowledge about immigration increased during the discussion. (Mean = 6.89)
3. My readiness to participate in societal activities increased during the discussion. (Mean = 5.94)
4. I would be happy to participate in a similar discussion again. (Mean = 7.98)
5. Other participants took my opinions into account during the discussion. (Mean = 8.05)
Swedish language (Mean = 8.47, SD = 1.14, Alpha = 0.64)
1. The relations between the participants remained good throughout the discussion. (Mean = 9.13)
2. Generally speaking, I identified with the other members in my group. (Mean = 7.54)
3. All discussants had equal opportunities to be heard, and nobody was left out of the discussion. (Mean = 8.46)
4. The opinions presented in our discussion were based on facts. (reversed) (Mean = 7.64)
5. Inappropriate remarks or name-calling occurred in the discussion. (reversed) (Mean = 9.36)
6. Some participants provoked others in the discussion. (reversed) (Mean = 8.71)

SD standard deviation, Alpha Cronbach's Alpha

taking part was a pleasant experience. The second asks whether they think their issue knowledge increased as a result of taking part in the event. This question taps into the evaluation of the epistemic benefits of the deliberation process. The third question relates to the participants' readiness to take part in civic activities after the deliberative event. The fourth statement measures the participants' willingness to take part in a similar event anew. The fifth measures the internal inclusion in the discussion, i.e. whether they found that they could easily put forward their views. Principal component factor analysis using a polychoric correlation matrix was executed to confirm that the items loaded on a single factor. Our *process satisfaction index* is obtained by calculating the arithmetic mean of the five responses. The Cronbach's Alpha scores are acceptable: 0.60 (nuclear power) and 0.61 (immigration).

In the third experiment, the set of process satisfaction questions was not identical with the first two experiments. Thus, we construct an alternative index to capture the participants' satisfaction with the deliberative process (see Table 4). The first question asks whether the relationships between the participants remained good during

the discussion. The second captures to what extent a participant related to the other members of the group. The third question is about inclusion in terms of each participant having had an equal opportunity to be heard. The fourth question asks whether arguments were perceived to be based on facts. The fifth and sixth questions concern respect: if inappropriate comments or name calling were observed or if somebody deliberately provoked others. As above, we construct our *process satisfaction index* by calculating the arithmetic mean of the six responses. Cronbach's Alpha is 0.64 and indicates that the scale has an acceptable internal consistency.

We perform three separate multivariate regression models to establish if and how group level and individual level disagreements are related to satisfaction with the deliberative experience. We use ordinary least squares (OLS) models with cluster-robust standard errors to allow for intragroup correlation. In other words, the observations are independent across groups but not necessarily within groups. This is appropriate since clusters of individuals were assigned to different treatments. We also control for the treatments by using dummy variables. In the nuclear power experiment, a dummy variable takes into account those groups where common statements were written (in contrast to a vote within the group). In the immigration experiment, one dummy variable controls for a con enclave and another dummy variable for a pro enclave (while the cross-cutting group is the reference category). In the Swedish language experiment, we have two dummy variables to account for pro enclave (pro = 1, con = 0) and the presence of deliberative rules and moderation (yes = 1, no = 0).

Results

We present the findings for each experiment in turn. For each experiment we have two regression models: one model which includes only the linear terms for the two disagreement measures and another model which includes both the linear and squared terms. The inclusion of the squared term is of particular interest because we seek to assess if there is a curvilinear relationship between disagreement and process satisfaction (as in Esterling et al., 2015).

The regression estimates in Tables 5, 6 and 7 show that both group level and individual level disagreement are weak or no predictors of satisfaction with deliberation. First the adjusted R squared values are low, ranging between zero and 0.070. Second, all but two of the regression coefficients are statistically insignificant. In two of the experiments—immigration and language policy—group level disagreement does not have any impact on how satisfied the participants are with the discussion. Only in the experiment on nuclear power are there statistically significant relationships: between the linear term for group level disagreement and process satisfaction as well as between the linear term for individual level disagreement and process satisfaction. This implies that the satisfaction with the deliberative experience became lower the more the participants' opinions deviated from each other before deliberation, and the more the individual disagrees with the group. The reasons why the adjusted R squared values are higher in the two models in Table 5, which are intended to explain process satisfaction with

Table 5 Deliberation on nuclear power: effects of disagreement on process satisfaction

	Linear model Est (SE)	Curvilinear model Est (SE)
Group level disagreement		
Linear term	-0.404 (0.100)**	-0.374 (0.206)
Squared term	-	-0.029 (0.192)
Individual level disagreement		
Linear term	-0.148 (0.054)*	-0.089 (0.045)
Squared term	-	-0.082 (0.033)
Treatment (ref. Vote)		
Common statement	-0.041 (0.242)	-0.117 (0.252)
Constant	8.392 (0.161)**	8.652 (0.220)**
N	135	135
Adjusted R squared	0.055	0.070

Standard errors adjusted for 12 clusters (discussion groups)

Asterisks indicate coefficient significance level (two-tailed): ** $p < 0.01$; * $p < 0.05$

Table 6 Deliberation on immigration: effects of disagreement on process satisfaction

	Linear model Est (SE)	Curvilinear model Est (SE)
Group level disagreement		
Linear term	0.248 (0.140)	0.156 (0.158)
Squared term	-	0.143 (0.093)
Individual level disagreement		
Linear term	0.135 (0.128)	0.172 (0.170)
Squared term	-	-0.027 (0.057)
Treatment (ref. Cross-cutting)		
Pro enclave	0.491 (0.273)	0.453 (0.255)
Con enclave	0.049 (0.303)	-0.019 (0.292)
Constant	7.249 (0.199)**	7.234 (0.198)**
N	207	207
Adjusted R squared	0.015	0.009

Standard errors adjusted for 26 clusters (discussion groups)

Asterisks indicate coefficient significance level (two-tailed): ** $p < 0.01$; * $p < 0.05$

deliberation on nuclear power, are: a significant linear relationship between group level disagreement and process satisfaction and; a distinguishable curvilinear, yet non-significant, relationship between individual level disagreement and process satisfaction.

What is important to stress is that the squared term is substantially and statistically significant in none of the regression models. In other words, a curvilinear

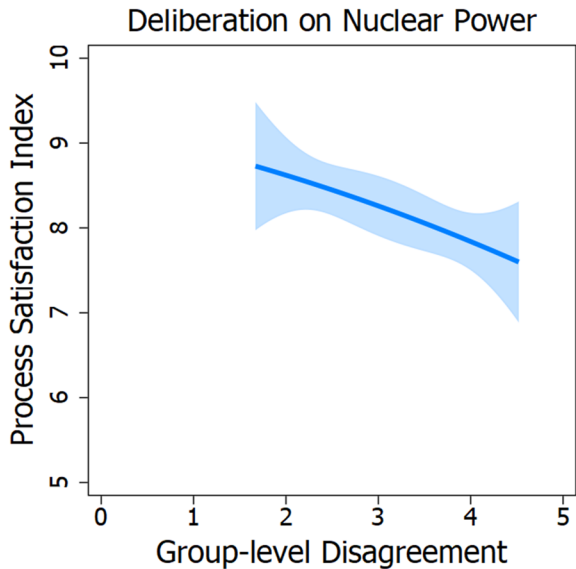
Table 7 Deliberation on the Swedish language: effects of disagreement on process satisfaction

	Linear model Est (SE)	Curvilinear model Est (SE)
Group level disagreement		
Linear term	-0.002 (0.161)	0.156 (0.306)
Squared term	-	-0.128 (0.145)
Individual level disagreement		
Linear term	-0.123 (0.101)	-0.096 (0.155)
Squared term	-	-0.011 (0.030)
Treatment 1 (ref. Con enclave)		
Pro enclave	-0.280 (0.198)	-0.272 (0.199)
Treatment 2 (ref. No rules)		
Rules	0.215 (0.200)	0.200 (0.208)
Constant	8.531 (0.183)**	8.571 (0.198)**
N	178	178
Adjusted R squared	0.007	0.000

Standard errors adjusted for 31 clusters (discussion groups)

Asterisks indicate coefficient significance level (two-tailed): **p < 0.01; *p < 0.05

Fig. 1 Predicted satisfaction with deliberative discussion on nuclear power in Finland by group-level disagreement. Predictive margins with 95% CI



relationship between disagreement and satisfaction with deliberative discussion cannot be found either at the group level or at the individual level. Hence, our results are not in line with the findings of Esterling et al. (2015), which identified that middle level disagreement within a group leads to the highest process satisfaction.

Fig. 2 Predicted satisfaction with deliberative discussion on nuclear power in Finland by individual-level disagreement. Predictive margins with 95% CI

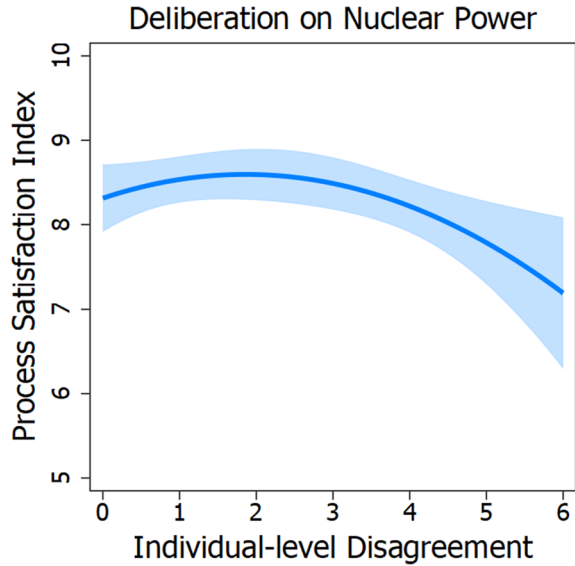
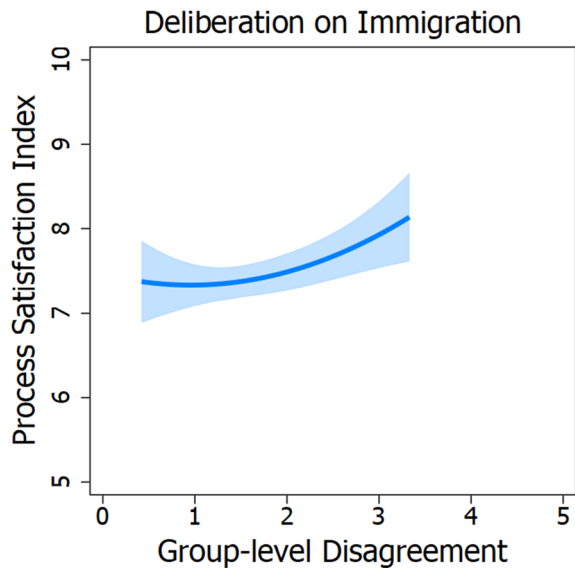


Fig. 3 Predicted satisfaction with deliberative discussion on immigration in Finland by group-level disagreement. Predictive margins with 95% CI



We also present predictive margins in addition to estimated regression coefficients. Predictive margins are estimates of the response mean when predictors are fixed at specified values. Our purpose is to visually show to what extent the mean response slopes, or curves, across deliberation experiments resemble each other. The predictive margins reported in Figs. 1, 2, 3, 4, 5 and 6 are based on the regression models with both the linear and squared terms for disagreement. Group level and

Fig. 4 Predicted satisfaction with deliberative discussion on immigration in Finland by individual-level disagreement. Predictive margins with 95% CI

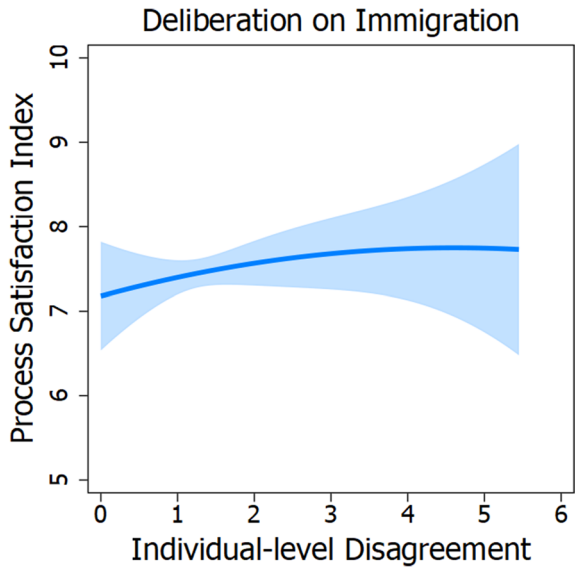
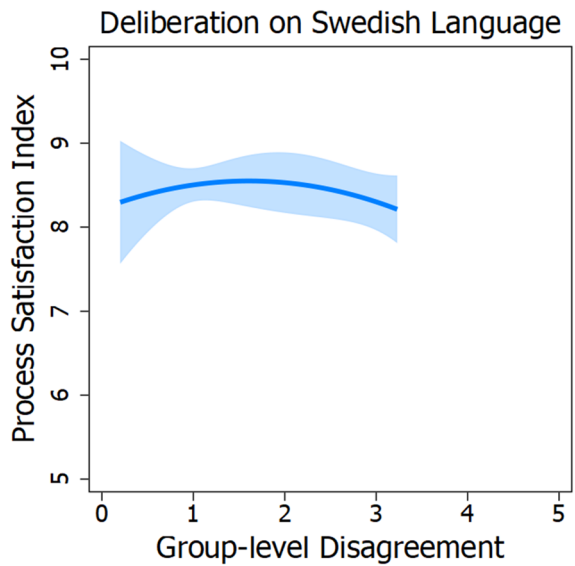
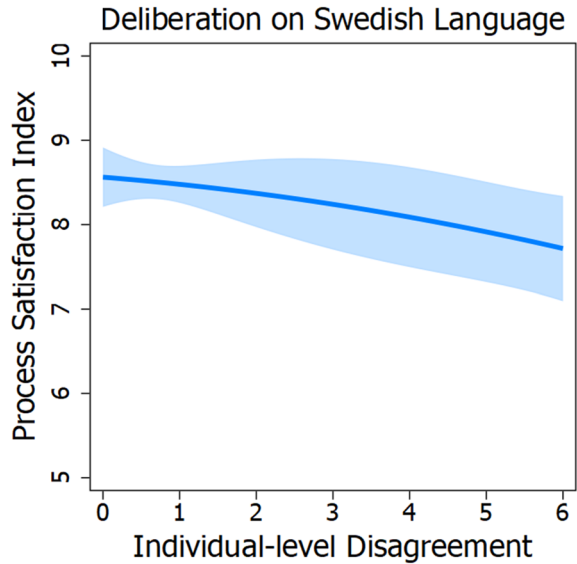


Fig. 5 Predicted satisfaction with deliberative on the role of the Swedish language in Finland by group-level disagreement. Predictive margins with 95% CI



individual level disagreement are fixed at evenly spaced values that range between the minimum and maximum values, separately for each experiment and disagreement index, while holding all other independent variables at their actual values. The predicted process satisfaction of each group and individual is then calculated based on the estimated regression coefficients. Finally, the average of the predicted value

Fig. 6 Predicted satisfaction with deliberative discussion on the role of the Swedish language in Finland by individual-level disagreement. Predictive margins with 95% CI



at the specified values for each group and individual in each experiment is calculated and presented.

If we begin by examining the effect of group level disagreement, the slopes/curves differ widely from each other. In the experiment on nuclear power, the relationship between group level disagreement and the process satisfaction index is negative and predominantly linear in form (see Fig. 1). As noted above, the regression coefficient for the linear term is statistically significant and the effect is substantive. According to Fig. 1, the process satisfaction index is about 8.7 if group level disagreement is set to its minimum (i.e. when the standard deviation of the arithmetic mean of attitudes toward nuclear power before deliberation is set to 1.7). When group level disagreement is set to its maximum (4.5), the process satisfaction index drops to almost 7.6 points. The slope of the curve for group level disagreement in turn is positive instead of negative in the experiment on immigration (Fig. 3), while the shape of the curve is slightly curvilinear in the experiment on the Swedish language (Fig. 5).

For individual-level disagreement, the slopes also differ. First, the curve is inversely U-shaped in the experiment on nuclear power (Fig. 2), which was anticipated to be the case. When individual level disagreement is set to 0 and 6, the predictive margins for process satisfaction are 8.3 and 7.2 correspondingly. However, the regression coefficients for the linear and squared terms are not statistically significant. Further, the curve is barely inversely U-shaped in the experiment on immigration, and the difference in predicted process satisfaction is minor when individual-level disagreement is set to its minimum and maximum values (Fig. 4). Finally, the slope is linear and negative, but since the confidence intervals are broad for the experiment on the Swedish language, we can deduce that when an individual's own opinions deviate largely from the group mean, then his or her satisfaction with deliberation is not lower than that of other participants (Fig. 6).

In the light of non-significant results, we perform post-hoc power analyses to ascertain if our study has enough large samples power to detect statistically significant effects. The question is whether the squared term for group level disagreement, which represents the curvilinear effect, could reach a given magnitude. To assess the expected effect, we make a theoretical prediction based on the shape and amplitude of the predicted curvilinear relationship between table disagreement and process satisfaction in the study by Esterling et al. (2015). Let us assume that group level disagreement varies between 0 and 3, and the outcome variable is 8 when disagreement is zero, 9 at the vertex of the parabola and 7 when disagreement is at its maximum. In that case, we expect the coefficient for the squared term to be -0.65 . This is a more conservative value compared to Esterling et al. because we estimate their coefficient to be about twice as large if we consider that the variables have different scales. Then we use the observed standard deviations for group level disagreement and process satisfaction in each experiment. The null hypothesis is that the coefficient is zero. All computations, based on a linear regression t-test on the slope of a simple linear regression, are performed for a two-sided hypothesis where the significance level is set to 0.05 and power to 90 percent. The minimum sample sizes required are 69, 117 and 111. These are considerably less than our samples of 135, 207 and 178 participants. We also determine the minimum values of the regression coefficients that can be detected with our observed samples at 90 percent power: -0.47 , -0.49 and -0.52 . We thus believe that our null results are reliable; they were unlikely to have occurred simply because we did not have enough participants.

When it comes to the controls for experimental treatments, none of them have a significant effect. First, process satisfaction is not affected by whether or not the group wrote a common statement on whether to build more nuclear power plants (Table 5). Second, there is no difference in process satisfaction between the groups that consisted of people with similar (either pro or con) or mixed opinions toward immigration (Table 6). Third, the two dummy variables which distinguish between pro and enclaves, on the one hand, and between enclaves with and without facilitation and rules, on the other hand, are insignificant as well. We can therefore conclude that the specific setting of deliberation, such as topic or experimental treatment, does not appear to influence process satisfaction in any systematic manner.

Conclusions

Our results show that people who participate in organized deliberation seem to be satisfied with the process in general. In a comparable study, Esterling et al. (2015) observed that participants in groups with medium levels of initial disagreement liked deliberation the most. However, in our case, we could not establish a similar curvilinear relationship between disagreement and satisfaction with deliberation. We could neither find strong evidence that participants in groups with a high level of internal disagreement, or individuals who radically deviated from the group mean, were much less satisfied with the experience. Only in the nuclear power experiment was there a small statistically significant *linear* association between disagreement and process satisfaction. This was found especially at the group level.

We have no clear answers why our results deviate from Esterling et al. (2015), but we can consider certain possibilities. Esterling et al. use data from *California Speaks*, which was connected to a real political process, whereas we focus on deliberative mini-publics that were organized as controlled experiments for research purposes. It is possible that participants in our case tolerate disagreement better, because they know that their discussions do not have an impact on real-world politics. In other words, differences in results may follow from engaging in *cold* (in our case) versus *hot* (in *California Speaks*) deliberation (Fung, 2007). However, it is noteworthy that even though *California Speaks* was connected to a real political process the impact was indirect, and participants may not therefore feel that they can directly influence the ultimate decisions. In addition to political impact, the topic of deliberation may matter, and one way to study the difference of hot and cold deliberation in future research is to vary topics for deliberation. Do they have an impact on the tolerance of disagreement?

One possible explanation could relate to the way participants were recruited, assuming that tolerance of disagreement is related to individual characteristics. Esterling et al. (2015) analyzed data with self-selected participants, whereas our participants were recruited through random sampling. Furthermore, in our case, respondents whose opinions were close to the mean value on the index variables measuring baseline views, were excluded in the second and third experiments. If these “moderates” were different in terms of tolerating disagreement, their exclusion might have impacted the results. People may indeed represent different types in relation to politics (Berelson et al., 1954, p. 322–323; Hibbing & Theiss-Morse, 2002, p. 81) and that may influence their exposure (Mutz, 2006) and tolerance of disagreement. However, when comparing participants in the second and third experiments with the public at large in Finland, there seems to be no difference in their willingness to discuss politics with people who disagree with them.⁷ Nevertheless, we think that more research on the role of personal characteristics and behavior in structured deliberation is needed.

There are also other differences between our data and *California Speaks* pertaining to the deliberative process. Our small groups, for example, gathered in separate rooms, whereas the *California Speaks* discussions took place at round tables in large halls. It is also possible that there are cultural differences between Finnish and American participants.

One may also ask how well we are able to generalize from our observations based on controlled lab-in-the-field experiments to deliberative mini-publics connected to real decision-making processes. As we argue above, the connection to a real-world process may not be that crucial, because the impact of mini-publics is rarely direct, and it appears that participants in our experiments took the discussions seriously,

⁷ A survey item: “I prefer to discuss politics and societal questions with like-minded people” was asked in the second and third experiment. In the second experiment, 70 percent of the participants disagreed with this statement, in the third experiment, 64 percent of the participants disagreed. In the Finnish National Election Study of 2011 (Borg & Grönlund, 2011) the same question was asked of the whole electorate, among which 66 percent disagreed with the statement.

one indication of this is that people changed their opinions due to the discussions (Grönlund et al., 2015; Setälä et al., 2010; Strandberg et al., 2019). An ultimate test of the impact of a mini-public's connection to the actual political process would require a direct comparison between two types of mini-publics, one connected to a political process and another not connected. Further research in other deliberative settings could also be used to test the robustness of our findings.

Our results may also suggest that just as the literature on enclaves and deliberation is inconclusive on effects, the empirical reality might vary from deliberation to deliberation. Maybe other factors influencing participants' experiences are at play than just the average level of disagreement in the deliberating groups. Some possible explanations can be the topic of deliberation, the moderator's activity, and other group level factors. Nonetheless, especially based on the second experiment (immigration), we can conclude that our results give support to the finding that people who take part in facilitated deliberation do not feel negatively about political disagreement (cf. Himmelroos et al., 2017). In general, participants in groups with high disagreement find deliberation under organized forms equally appealing as participants in groups with low disagreement. Further, this result was robust against the variations in the analyzed deliberative settings, i.e. the decision-making method in the small groups, the composition of groups, the presence of an expert panel or even the presence of a moderator and rules of discussion.

Further, looking at the individual level, we find that the participants whose views differ a lot from the mean opinion of their group, do not find deliberation that much less appealing. Their process satisfaction was somewhat lower in two of the experiments, but not decisively so. In the experiments on nuclear power and the Swedish language, extreme "deviants" had, on a scale from zero to ten, only about a one point lower satisfaction level than participants who were close to the group mean. Considering that these are extreme cases, the effect should be considered weak.

Our results are in many ways positive news for deliberative theory in general and the use of deliberative mini-publics in democratic decision-making in particular. Creating a safe and facilitated environment, where lay citizens can gather and deliberate on political issues does not appear to lead to excessive negative experiences, even though the deliberating groups would have high levels of disagreement. Neither do people whose views on the discussed matter differ greatly from the group's mean opinion evaluate the deliberative process more negatively than others. This finding is also evident when we look at the support of using deliberative mini-publics in democratic decision-making. People who have participated in deliberation seem to be equally supportive of deliberative forums, no matter how much agreement or disagreement they experienced in the discussions of their group.

Acknowledgements We would like to thank all the colleagues who have contributed to the three experiments, as well as the anonymous referees and participants of conferences who have provided valuable comments on the paper.

Funding Open access funding provided by Åbo Akademi University (ABO). The experiments and research conducted here have been funded by Åbo Akademi Foundation, Centers of Excellence Funding, and the Academy of Finland, project number 251222, Academy research fellow project number

28200170, and the Strategic Research Council of the Academy of Finland project numbers 312671, 312675, 326662.

Data Availability Grönlund, Kimmo (Åbo Akademi University) & Herne, Kaisa (University of Turku) & Setälä, Maija (University of Turku): *Deliberative Discussion on Nuclear Power 2006* [dataset]. Version 2.0 (2018-07-23). Finnish Social Science Data Archive [distributor]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD2355>. Grönlund, Kimmo (Åbo Akademi University): *Citizen Deliberation on Immigration: Survey Data 2012* [dataset]. Version 1.0 (2014–11-05). Finnish Social Science Data Archive [distributor]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD2958>. Grönlund, Kimmo (Åbo Akademi University) & Herne, Kaisa (University of Tampere) & Himmelroos, Staffan (Åbo Akademi University) & Strandberg, Kim (Åbo Akademi University) & Setälä, Maija (University of Turku) & Bäck, Maria (Åbo Akademi University) & Leino, Mikko (Åbo Akademi University): *Citizen Deliberation on the Swedish Language and Swedish-speaking Minority in Finland: Recruitment Survey 2014* [dataset]. Version 2.0 (2018–07-17). Finnish Social Science Data Archive [distributor]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD3100>. Grönlund, Kimmo (Åbo Akademi University) & Herne, Kaisa (University of Tampere) & Himmelroos, Staffan (Åbo Akademi University) & Strandberg, Kim (Åbo Akademi University) & Setälä, Maija (University of Turku) & Bäck, Maria (Åbo Akademi University) & Leino, Mikko (Åbo Akademi University): *Citizen Deliberation on the Swedish Language and Swedish-speaking Minority in Finland: Post-deliberation Survey 2014* [dataset]. Version 1.0 (2016–07-15). Finnish Social Science Data Archive [distributor]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD3102>.

Code Availability All replication files (data and Stata Do files) are available from the Harvard Dataverse: <https://doi.org/10.7910/DVN/PG2RCO>

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval The studies were conducted in accordance with ethical standards.

Consent to Participate Consent to participate was obtained from all participants in the studies.

Consent for Publication Consent regarding publishing data gathered from the studies was obtained from all participants. All individual-level data have been anonymized.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdullah, C., Karpowitz, C., & Raphael, C. (2016). Affinity groups, enclave deliberation and equity. *Journal of Public Deliberation*. <https://doi.org/10.16997/jdd.258>
- Andersen, V. N., & Hansen, K. M. (2007). How deliberation makes better citizens: The Danish deliberative poll on the Euro. *European Journal of Political Research*, 46, 531–556. <https://doi.org/10.1111/j.1475-6765.2007.00699.x>

- Berelson, B. R., Lazarsfeld, P. F., & McPhee, W. N. (1954). *Voting. A study of opinion formation in a presidential campaign*. The University of Chicago Press.
- Borg, S., & Grönlund, K. (2011). Finnish National Election Study. Version 4.0 (2020-10-28). Finnish Social Science Data Archive [distributor]. Retrieved from <http://urn.fi/urn:nbn:fi:fsd:T-FSD2653>
- Dryzek, J. S. (2000). *Deliberative democracy and beyond: Liberals, critics, contestations*. Oxford University Press.
- Elster, J. (1998). *Deliberative democracy*. Cambridge University Press.
- Esterling, K. M., Fung, A., & Lee, T. (2015). How much disagreement is good for democratic deliberation? *Political Communication*, 32(4), 529–551. <https://doi.org/10.1080/10584609.2014.969466>
- Fung, A. (2007). Minipublics: Deliberative designs and their consequences. In S. W. Rosenberg (Ed.), *Deliberation, participation and democracy. Can the people govern?* (pp. 159–183). Palgrave Macmillan.
- Grönlund, K., Herne, K., & Setälä, M. (2015). Does enclave deliberation polarize opinions? *Political Behavior*, 37, 995–1020. <https://doi.org/10.1007/s11109-015-9304-x>
- Grönlund, K., Herne, K., & Setälä, M. (2017). Empathy in a citizen deliberation experiment. *Scandinavian Political Studies*, 40(4), 457–480. <https://doi.org/10.1111/1467-9477.12103>
- Gutmann, A., & Thompson, D. (1996). *Democracy and disagreement*. Belknap Press.
- Hibbing, J. R., & Theiss-Morse, E. (2002). *Stealth democracy. Americans' beliefs about how democracy should work*. Cambridge University Press.
- Himmelroos, S., Rapeli, L., & Grönlund, K. (2017). Talking with like-minded people—Equality and efficacy in enclave deliberation. *The Social Science Journal*, 54, 148–158. <https://doi.org/10.1016/j.soscj.2016.10.006>
- Karpowitz, C. F., Raphael, C., & Hammond, A. S. (2009). Deliberative democracy and inequality: Two cheers for enclave deliberation among the disempowered. *Politics & Society*, 37(4), 576–615. <https://doi.org/10.1177/0032329209349226>
- Kwak, N., Williams, A. E., Wang, X., & Lee, H. (2005). Talking politics and engaging politics: An examination of the interactive relationships between structural features of political talk and discussion engagement. *Communication Research*, 32(1), 87–111. <https://doi.org/10.1177/0093650204271400>
- Mansbridge, J., Bohman, J., Chambers, S., Christiano, T., Fung, A., Parkinson, J., & Warren, M. (2012). A systemic approach to deliberative democracy. In J. Parkinson & J. Mansbridge (Eds.), *Deliberative systems: Deliberative democracy at the large scale* (pp. 1–26). Cambridge University Press.
- McClurg, S. D. (2006). Political disagreement in context: The conditional effect of neighborhood context, disagreement and political talk on electoral participation. *Political Behavior*, 28(4), 349–366. <https://doi.org/10.1007/s11109-006-9015-4>
- Mercier, H., & Landemore, H. (2012). Reasoning is for arguing: Understanding the successes and failures of deliberation. *Political Psychology*, 33(2), 243–258. <https://doi.org/10.1111/j.1467-9221.2012.00873.x>
- Morrell, M. E. (2010). *Empathy and democracy*. Penn State University Press.
- Mutz, D. C. (2002). Cross-cutting social networks: testing democratic theory in practice. *American Political Science Review*, 96, 111–126.
- Mutz, D. C. (2006). *Hearing the other side. Deliberative versus participatory democracy*. Cambridge University Press.
- Pattie, C. J., & Johnston, R. J. (2009). Conversation, disagreement and political participation. *Political Behavior*, 31(2), 261–285. <https://doi.org/10.1007/s11109-008-9071-z>
- Setälä, M., Grönlund, K., & Herne, K. (2010). Citizen deliberation on nuclear power: A comparison of two decision-making methods. *Political Studies*, 58, 688–714. <https://doi.org/10.1111/j.1467-9248.2010.00822.x>
- Strandberg, K., Himmelroos, S., & Grönlund, K. (2019). Do discussions in like-minded groups necessarily lead to more extreme opinions? Deliberative democracy and group polarization. *International Political Science Review*, 40(1), 41–57. <https://doi.org/10.1177/0192512117692136>
- Sunstein, C. (2002). The law of group polarization. *The Journal of Political Philosophy*, 10(2), 175–195.
- Theiss-Morse, E., & Hibbing, J. R. (2005). Citizenship and civic engagement. *Annual Review of Political Science*, 8, 227–249. <https://doi.org/10.1146/annurev.polisci.8.082103.104829>