

# Varieties of Opinion Change in a Finnish Citizen Deliberation Experiment on Immigration

Marina Lindell (Åbo Akademi University), André Bächtiger (University of Stuttgart), Kimmo Grönlund (Åbo Akademi University), Kaisa Herne (University of Tampere), Maija Setälä (University of Turku)

## Abstract

In the study of deliberation, a largely underexplored area is why some participants become more extreme, whereas some become more moderate. Opinion polarization is usually considered a suspicious outcome of deliberation, while moderation is seen as a desirable one. This article takes issue with this view. Results from a deliberative experiment on immigration show that polarizers and moderators were not different in their socio-economic, cognitive, or affective profiles. Moreover, both polarization and moderation can entail deliberatively desired pathways: in the experiment, both polarizers and moderators learned during deliberation, levels of empathy were fairly high on both sides, and group pressures barely mattered. Finally, the absence of a participant with an immigrant background in a group was associated with polarization in anti-immigrant direction, bolstering longstanding claims regarding the importance of *presence* in interaction (Philips 1995).

Paper presented at the “13ème Congrès National Association Française de Science Politique (AFSP), Aix-en-Provence, June 22–24, 2015

## Introduction

Empirical studies of citizen deliberation suggest that participants often change opinions (and also quite radically; see, e.g., Fishkin, 2009). Luskin et al. (2002) claim that knowledge gain is an important mechanism of opinion change, whereas Sanders (2012) was unable to identify any robust predictor of opinion change in a recent study based on a pan-European deliberative poll (Europolis). A largely understudied area in this regard is why some participants polarize their opinions due to deliberation, and why others moderate them. Moderation is normally seen as a desirable outcome of a deliberative process: by carefully listening to others, participants with extreme opinions realize that there is merit in other's positions and arguments. Polarization, by contrast, is frequently considered as a suspicious outcome. According to Sunstein (2002), group polarization reflects a dynamic psychological process, whereby groups move to the extreme on the basis of biased information processing and biases in the argument pool. Recent lines of theorizing put a question mark on this interpretation, arguing that polarization may not necessarily be a bad thing since it may simply reflect preference clarification, that is, people become aware of what they really want (Knight and Johnson, 2011). In other words, the reinforcement of existing opinions may have deliberative dimensions (or, is at least not anti-deliberative).

Despite some recent contributions on polarization and moderation (Sunstein 2009, Jones 2013, reference omitted) we still know very little what explains these tendencies at the individual level. This paper addresses this gap in the literature by focusing on participants who have changed their minds more than average in a deliberative event, either in a more extreme or more moderate direction.

The paper explores what drives polarization and moderation by focusing on a batch of individual-level and group-related variables. In concrete, we focus on education, knowledge on the topic, empathy, understanding and group pressures, and test how they affect polarization and moderation. Furthermore, we engage in a normative assessment of polarization and moderation: the cognitive, affective, and group-related factors that we explore also carry normative significance, and we evaluate whether the moderation or polarization of opinions occurs in accordance with normatively desired pathways or not.

We assess polarization and moderation in the context of an experiment where 207 citizens deliberated on immigration in Turku (Finland). This experiment is particularly useful for our research purpose, since it manipulated the group context, by assigning participants in like-minded groups ( ) and diverse opinion groups (mixed groups). The experiment was designed in order to test what happens when like-minded opinion groups deliberate with deliberative discussion rules and the presence of a trained moderator.

Even though group polarization did not occur to a large extent (reference omitted), both polarization and moderation are observed at the individual level.

The structure of the paper is as follows. First, we provide a theoretical discussion on polarization and moderation of opinions, identifying its individual-level and group-related antecedents. Second, we present the procedures of the experiment. Third, we report the results, followed by a discussion. Concluding remarks are presented at the end of the paper.

### **The polarization and moderation of opinions: a theoretical framework**

Theories of deliberative democracy contain many empirical assumptions about preference and opinion formation. The thrust of deliberative approaches is that citizens are often not adequately informed about political issues and have not sufficiently engaged in weighing reasons for different policy positions (Fishkin 2009; Muhlberger & Weber 2006). Existing empirical findings are mixed in terms of the extent of opinion change in the aggregate. Some studies show major and radical changes in opinions at group level (Fishkin & Luskin, 1999; Luskin et al., 2002; Goodin & Niemeyer, 2003; Blais et al., 2008) whereas others show only minor changes (Denver et al., 1995; Merkle, 1996; Hall et al., 2011). But most of the time, there are opinion changes at the *individual level* with movements in different directions that go undetected at the group level (Barabas, 2004; Andersen & Hansen, 2007). To understand the dynamics of opinion formation and opinion change, it is necessary to better understand individual level processes underlying opinion change. In particular, why do some participants polarize their opinions, while others moderate them? This is not only interesting from an analytical and empirical point of view, but also carries significant normative meaning.

To date, the concept of polarization is mainly used at the aggregate level. According to Sunstein (2002: 176) "*group polarization means that members of a deliberating group predictably move toward a more extreme point in the direction indicated by the members' pre-deliberation tendencies*". In this paper we elaborate the polarization and moderation at an individual-level. Polarization is used to indicate that individuals move towards the extreme by strengthening their initial opinion (see, e.g., Wojcieszak, 2011 for a similar interpretation). Reflecting on polarization studies in psychology, Miller et al. strongly emphasize "the importance of considering individual differences in the direction of reported attitude change."

The standard assumption in the group polarization literature is that the polarization of opinions is due to group context (an aspect that we will discuss in more detailed fashion further below). Sunstein (2002) has repeatedly argued that if participants enter deliberation with only likeminded people they

become more extreme in face-to-face deliberation. Biased assimilation of information and biases in the argument pool are assumed to be the underlying causes for opinions becoming more extreme. More recently, however, the literature has offered contradictory findings about deliberation in enclaves. Findings from Karpowitz et al. (2009) and [reference withheld] contradict the polarization hypothesis of enclave deliberation since they found no clear evidence of group polarization and amplification of cognitive errors when likeminded groups deliberated with moderators and rules.

By contrast, many deliberative scholars consider moderation as an ideal outcome of a deliberative process. O'Flynn (2006), for instance, argues that moderation represents a marker of political equality: "if we enter into a democratic process in a spirit that recognizes that other citizens have equal standing with ourselves, we shall be ready to moderate our claims, since this is what equality requires in the face of the different and competing views of our fellow citizens." (p. 738) From an epistemic perspective, some have argued that epistemic peers faced with disagreement should be "conciliatory" (e.g., Feldman 2007; Christensen 2007). Empirically, a properly set-up deliberative process where people are exposed to information and arguments that differ from their initial understanding might open up for (even strong) opinions to change and help participants to reach consensus and enlightenment (Barabas, 2004: 689–690). Individuals with strong opinions may also become more uncertain and ambivalent when they interact with others and get to know the 'other side' (Mutz, 2006: 102–105; Ackerman and Fishkin, 2004: 53). Indeed, there are several studies indicating that opinions converge or move to the middle after deliberation (Farrar et al., 2010; Lang, 2007; Schweigert, 2010; French and Laver, 2009).

In recent years, however, deliberative theory has opened itself up to a broader variety of opinion transformations, going beyond moderation and the search for consensual solutions. For instance, participants in a deliberative process might initially think that their preferences are reconcilable (or not too distant), but find out in discussion that the opposite is actually true. Some epistemic scholars also claim that that conciliation is a problematic strategy. Not only are there multiple reasonable perspectives on an issue (Cohen 1995), the available evidence frequently underdetermines the choice among competing theories, i.e., more than one theory is supported by the facts at hand. Using Bayesian network analysis, Jern et al. (2011) show that some instances of polarization are consistent with what they call "a normative account of belief revision". For instance, it is entirely possible that when two persons see the same piece of evidence, different webs of beliefs may be conducive to polarization rather than consensus. Benoit and Dubra (2014) concur: "when a person is presented with equivocal evidence, that is, evidence that can reasonably interpreted as being either in favour or against a proposition, his beliefs can reasonably move either towards or away from accepting the proposition, or not move at all, and by that very fact, the harmonization, moderation, and polarization

of two individuals are all reasonable outcomes.” In a similar vein, Knight and Johnson (2011) argue: “even if as a result of the increased information that political argument makes available, individuals come to hold their preferences more reflectively, it in no way follows that this will lead to greater substantive agreement at the aggregate level.” (p. 145) Knight and Johnson (2011) consider clarification and “structured disagreement” more important than opinion change *per se*; and clarification may well encompass polarization, moderation, or stability of opinions. In accordance with Sunstein (2006), however, deliberative democrats would insist that if polarization (or, moderation) occurs, it should do so in normatively defensible ways and not be the product of undesirable group dynamics or on other non-deliberative pathways. We shall return to this issue in due course.

Below, we discuss a number of individual-level and group-related factors that can be assumed to affect opinion change in general and the polarization and moderation of opinions in particular. There are, of course, myriads of factors, which may drive the polarization and moderation of opinions; we shall concentrate on those factors that are not only particularly relevant from an empirical-analytical but also from a normative perspective. In concrete, we focus on cognitive and socio-economic variables, empathy and trust, group effects, and ideology.

### *Individual-level factors*

#### *Cognitive and socio-economic variables*

The general assumption is that higher levels of education and knowledge make more sophisticated deliberators: better reasoning skills and higher knowledge levels involves a better ability to consider reasoned arguments of others (Rosenberg 2007, 342-343). With regard to the polarization and moderation of opinions, one hypothesis is that high education and high knowledge on the issue fosters the moderation of opinions, for two reasons. First, more educated and more knowledgeable persons may have higher democratic values leading to higher levels of tolerance and respect and hence the moderation of opinions. Second, more educated and more knowledgeable persons may also possess higher levels of cognitive complexity (Suedfeld, et al., 1992; Tetlock, 1983). Cognitive complexity captures the degree to which an individual perceives, distinguishes and integrates various dimensions of an issue under discussion. High scores of cognitive complexity indicate that an individual is able to accommodate conflicting goals or values (Gruenfeld and Preston 2000) as well as to recognize that different opinions are legitimate and can be held simultaneously. Consequently, one can expect more educated and more knowledgeable persons with higher levels of cognitive complexity to moderate their opinions when they are confronted with other viewpoints and counterarguments, whereas as

persons with lower education, lower knowledge and lower levels of cognitive complexity to be more conducive to polarization. However, a counter-hypothesis derived from Zaller suggests that highly educated and knowledgeable persons more strongly hold onto their attitudes. Thus they may be more resistant both to moderation and polarization trends than individuals with low education and low knowledge who are not so confident in their opinions and thus find it easier to assimilate new information and subsequently change their minds. Barabas (2004) found that participants with high knowledge who debated social security policy changed their views the least, and, when deliberation was not consensual, they even strengthened their prior opinions.

Other socio-demographic factors, such as gender and age, might also be influential in shaping opinion formation. However, empirical findings are contradictory when it comes to the impact of gender and age on opinion change. In a citizen deliberation in Finland year 2006 the oldest participants changed their opinions the most (towards lower support for nuclear power), while education had no effect on opinion change (Setälä et al., 2010). French and Laver (2009) experiment found that participants with low education changed their opinion the most while age had no effect. Some studies indicate that women are more easily persuadable than men and respond more to debate and influence (reference). However, much seems to depend on the topic of deliberation since the gender effect can be reversed for topics where women have strong attitudes (Suiter et al. 2014, 3). Overall, it seems difficult to make clear-cut predictions for socio-economic factors and polarization or moderation trends. We leave it to the empirical analysis to explore relevant effects.

### *Empathy and Trust*

The term empathy has been defined in a number of ways and various elements have been connected to empathy (e.g. Hoffman 2000; Preston and de Waal 2002; Walter 2012). While some writers emphasize the affective element of empathy, i.e. reproducing others' sentiments, others separate affective from cognitive empathy, i.e. understanding others' perspectives (Walter 2012).

Independent of the specific characterization of empathy it is clearly a relation between a subject who experiences empathy and a target who gives rise to empathy. Roughly we can say that empathy is the ability to put oneself in the other's position. Empathy is therefore characterized as a method for gaining knowledge about the experiences and internal states of others by imagination or internal simulation and reproduction of those states. Many theorists of deliberative democracy emphasize the cognitive side of empathy (e.g., Chambers 1996: 100). Perspective taking leads people to understand what others think and thus increases the likelihood that people have less negative evaluations of out-groups.

Mutz (2002:121) concurs that people with higher predispositions for perspective taking have higher level of tolerance when exposed to dissonant views. This makes them more open to opinions that conflict with their own. Indeed, if participants engage in perspective taking, then polarization tends to decrease (Morrell 2010:114,126). The emotional side of empathy is about role-taking, i.e. to familiarize oneself with other's situation and to feel what they feel (Morrell, 2010:58-62). Research suggests that emotions might play both positive and negative roles in a deliberative process. Emotions can help citizens to filter new information and to engage with others but emotions can also contribute to citizens disengaging from public life (Delli Carpini et al. 2004:328). It seems important that participants get a chance to become emotionally involved and to empathize with one another. Emotional impetus is needed for participants to reason together about the common good (Rosenberg 2007:348). If there is a lack of empathy, existing cleavages may be reinforced causing opinions to polarize (Rosenberg 2007:355). Instructing people to pay attention to others' feelings might decrease the tendency to attribute negative motives to people with opposite views and increase positive perceptions of out-groups (Morrell 2010:107,114). Hence, both cognitive and emotional aspects of empathy converge in predicting the moderation of opinions. Since the cognitive and emotional aspects of empathy are strongly correlated, we analyze empathy as a compound phenomenon in the empirical analysis.

Empathy is closely related to the concept of social trust. With social trust we mean generalized interpersonal trust as opposed to particularized interpersonal trust. According to Newton (2007:344-345) particularized interpersonal trust means trusting only individuals you know (i.e. trust my family or my social group), while generalized interpersonal trust entails general feelings that most people can be trusted. Generalized interpersonal trust enhances the feeling of a shared identity. Responses in surveys about generalized interpersonal trust can be taken as a measure of how people treat each other and how they experience their social relations (Newton 2007:344-349). According to Uslaner (2001), levels of trust do not depend on temporary experiences, i.e., attitudes towards other people are quite stable and the experiences gained by interacting with others do not affect the level of trust significantly. Trust is said to contribute to social integration, co-operation, personal life satisfaction, and optimism (Uslaner 2001). Both empathy and trust are closely related to a good understanding of others' opinions. We expect high values of generalized interpersonal trust to induce cooperative attitudes in the face of disagreement, and thus lead high trusting people to moderate their opinions.

### *Group Effects*

As mentioned above, opinion polarization and moderation can be affected by the composition of groups. Deliberation in like-minded groups – or, enclaves - can breed groupthink since the argument

pool might be limited and the group makes poor decisions based on incomplete and biased information. By contrast, in groups with opinion diversity, groupthink mechanisms should not set in, thus stifling opinion polarization.

Two types of mechanisms behind polarization have been identified, namely social comparison and persuasive arguments (Farrar et al. 2009, 616; Isenberg 1986; Sunstein 2002, 179-180). Social comparison refers to the tendency of individuals to act in order to win social acceptance from other members of the group. In order to be accepted, individuals need to process information of how other people present themselves, and adjust their own behavior accordingly (Isenberg 1986, 1142). Individuals may act in different ways in order to be perceived favorably by other group members. First, they may try to adjust their opinions according to the dominant view in the group. Social psychological experiments have also demonstrated that group pressures work in the way that people tend to conform to the views of the majority (Asch 1948). If someone agrees with you, you're apt to like that person more and since everybody wants to be liked, this imposes a lot of pressure on people with views inconsistent with the group's consensus (Sunstein 2006: 68).

The other mechanism behind group polarization, namely the deployment of persuasive arguments, is based on the idea that individuals are convinced by the contents of arguments put forward in the group. Consequently, if arguments heard in a group are biased in one direction, there is likely to be a further shift to this direction. Group polarization is likely to be reinforced by biases in information processing. 'Confirmation bias' is a well-established phenomenon, which means that people are inclined to seek information confirming their prior beliefs and to disregard information against them (Mercier and Landemore 2012, 251). More generally, motivated reasoning refers to a variety of cognitive and affective mechanisms, which lead individuals to arrive at the conclusions they want to arrive at (Kunda 1990). In a group of like-minded people, individual biases in information processing and reasoning are not checked by arguments put forward by individuals supporting conflicting views. Opinions are likely to polarize because individuals only hear arguments supporting their own prior position – in fact; they may even hear new arguments in support of it.

With regard to polarization and moderation, we focus on the effects of group pressure. One plausible expectation is that group pressure might be conducive both to the polarization *and* moderation of opinions: in enclave groups with like-minded people, participants might feel pressure to go the 'extremes', whereas in non-enclave groups with diverse opinions, participants might also experience group pressure to go to the "middle". However, group dynamics might also go fully unnoticed by participants, especially in enclave groups. By asking participants about their discussion experiences, we make a first stab to shed some light on these questions.



Finally, another group composition effect, namely the presence of the *other*, might affect opinion formation as well. There are long-standing claims in the literature that the *physical presence* of less privileged or marginalized groups is not only a democratic predicament (Philips 1995) but matters for outcomes as well. According to social identity theory members of a group might have a tendency to emphasize their similarities, i.e. strengthening in-group identity, and thus seeking to find negative aspects of out-groups. The physical presence of out-groups may be an important factor to reduce such tendencies (Hogg, 1993).

### *Ideology*

Since we take the direction of opinion change into account, namely the polarization and moderation towards pro- and anti-immigration directions, we also need to consider ideological factors. Modern cleavage literature suggests the existence of a two-dimensional map of economic and cultural integration or demarcation (see Kriesi et al., 2006). Consequently we focus both on left-right ideology and openness/closedness towards immigrants. We expect that persons with clear ideological profiles are more confident about their positions and thus might either keep their opinions on immigration or polarize in one or the other direction. Consequently, a clear leftist and openist or progressive stance may facilitate post-deliberative polarization in pro-immigration directions, whereas a clear rightist and conservative stance facilitates post-deliberative polarization in anti-immigration directions.

### *Normative considerations*

Finally, our study goes beyond a simple inventory of the drivers behind the polarization and moderation of opinions. The factors we introduced above also carry important normative significance. Assume that we find that opinion polarization was based on low knowledge, low levels of empathy, and on the basis of group pressures. We then would conclude that the polarization of opinion is indeed an undesirable outcome. But if the converse were true, namely that opinion polarization is associated with high knowledge, high levels of empathy, and no group pressures, we would conclude that polarization, even though considered as a suspicious outcome by many deliberative democrats, is at least not anti-deliberative. Or, similarly, the moderation of opinions even though considered as generally desirable by many deliberative democrats becomes questionable when it is based on low knowledge, low levels of empathy, and group pressures. We will normatively judge the polarization and moderation of opinions on the basis of four factors (see also Sanders 2012; Baccaro et al., 2015): epistemic advancement and capacity (knowledge and education); empathy and understanding, which

closely relate to the ethical dimension of deliberation (see Mansbridge et al., 2012); group pressures (both factual and perceived), and (perceived) 'deliberativeness' of the discussion process. While the first three factors have been discussed before, a word on the fourth factor – (perceived) 'deliberativeness' - is in order. By focusing on perceived 'deliberativeness', we try to capture whether participants found the discussion process useful, productive, and in accordance with ethical deliberative principles (such as civil and inclusive discussion). If participants rate the process 'deliberative' according to these standards, polarization (or, moderation) trends – in conjunction with the presence of epistemic advancement, empathy and understanding, as well as the absence of group pressures - take on deliberative dimensions. Under such conditions, polarization (or, moderation) trends can also be seen in accordance with the clarification function stressed by Knight and Johnson (2011).

Besides such cognitive, ethical, and group-related criteria, there may also be substantive considerations. When it comes to issues with a humanist or humanitarian dimension (such as immigration), Neblo (2007) shows that there is a tradition in deliberative theory holding that deliberation should be conducive to an expanded sense of community. Neblo (2007: 548) calls this "progressive vanguardism": "On this understanding, deliberative democracy is intrinsically and primarily an emancipatory project with strong substantive content, more or less tracking leftist political concerns." In concrete, this means that positions should move toward more progressive positions after deliberation (here: pro-immigration), whereas participants who already possess progressive positions should accentuate (or, keep) these positions. Clearly, "progressive vanguardism" is a highly contestable position, and many deliberative theorists are wary of making a statement regarding the directionality of opinion change. Indeed, a communitarian variant of deliberation argues that "good reasons" cannot be equated with liberal and progressive ideas and concepts even if these frequently reflect the dominant position among philosophers, but arise on the basis of communal values and self-understandings that mirror local and temporal circumstances (Forst, 2001). Therefore, our prime standard of normative evaluation will concern cognitive, ethical, and group-related factors.

## **Data**

### *The Experiment*

The topic of the deliberation experiment in Turku 2012 was immigration, a salient political issue in Finland. The purpose of the experiment was to compare deliberation processes and effects in two settings: 1) deliberation in likeminded groups (i.e. enclaves) and 2) deliberation in groups consisting of

people with clearly different opinions. Based on their initial opinions, the respondents were first classified as anti- or pro-immigrant i.e. assigned to two enclaves. After that, the participants were randomly allocated within their enclaves to likeminded groups, mixed groups and a control group.

A short survey (T1) was sent to a random sample of 12,000 adults in the Turku region. 39% responded to the survey of 14 questions that measured immigration attitudes. Those whose value for the sum variable was  $>8.3$  were included in the pro-immigration enclave and those whose value was  $<6.7$  were included in the anti-immigration enclave<sup>1</sup>. The second survey (T2) was sent to 2,601 persons. This survey also included an invitation to take part in a discussion about immigration. A gift certificate of 90 euros was offered to each participant in the event. 805 people volunteered and 366 were invited to take part in the deliberation; 207 people eventually showed up, with some bias towards the pro-immigration camp. The research team formed ten pro-immigration groups, five anti-immigration groups and eleven mixed groups. Immigrants were not excluded from the random sample and there were immigrants both in the pro and con groups.

Participants took part either on Saturday 31 March or Sunday 1 April 2012. The day started with a quiz measuring immigration related knowledge and general political knowledge (T3). This was followed by a short briefing event containing unbiased and basic facts about immigration; then the small group discussions began. Every group had a facilitator and discussed for four hours. The rules emphasized respect for other's opinions, the importance of justifying one's opinions and keeping an open mind towards other arguments and positions. The deliberation day ended with a survey (T4) repeating questions in T1, T2 and T3 as well as questions about how participants experienced the event.

## **Operationalization**

*Dependent variable: opinion change*

The dependent variable was formed from the responses of those participants who changed their minds on immigration more than average between T1 (before deliberation) and T4 (after deliberation), either in more extreme or more moderate directions. The variable is based on a sum variable of 14 questions (listed in the online Appendix, Cronbach's  $\alpha=0.94$ ). Each question was recoded into a scale from 0 to 1, whereby 1 indicates the most pro-immigration opinion. .

In order to identify the participants whose opinions did not change according to the mean change within their discussion group, we used a multistage process. As mentioned before, individuals with a

---

<sup>1</sup> Due to the experimental design only people with clear views about immigration were included, moderates with scores between 6.7-8.3 were excluded (n=631)

clear positive (pro-immigration) or negative (anti-immigration) view about immigration were initially included in the experiment. They were randomly assigned either to a like-minded treatment, where discussion took place in groups whose initial views were similar on the topic of immigration or a mixed treatment, where discussion took place in groups with an equal number of participants from both opinion enclaves (four each). We thereby obtain four experimental groups: pro-immigrants in like-minded group, pro-immigrants in mixed group, anti-immigrants in like-minded groups and anti-immigrants in mixed groups. Based on the aggregate opinion change in these four experimental groups we identified participants whose opinions did not change according to the mean change within the equivalent group. The standard deviations for opinion change in the groups range from 1.01-1.49. Frequently two or even three standard deviations are used in the literature as a measure to identify outliers. In this paper, however, we use only one standard deviation. First, using two standard deviations would leave us with a total of only 9 individuals, invalidating any meaningful statistical analysis. Second, by focusing on one standard deviation, we do not only capture the extreme outliers, but also atypical cases that do not fall within the normal distribution. Consequently, we focus on individuals who have changed one standard deviation or more than the “average opinion changers” in their experimental group. Over 30% of the participants have changed their opinion more than the average. We call them “big opinion changers”, whereby the term “big” does not refer to the absolute magnitude of opinion change, but to its deviance from the “average opinion changers” in the experimental groups. If the initial opinion was strengthened with one standard deviation or more the opinion change is considered as polarization. If the initial opinion was changed in the opposite direction (towards “the other side”) with one standard deviation or more the change is considered as moderation. 17%, or 35 individuals, have polarized their opinions, i.e., have reinforced opinions in the same direction as their initial opinion. 15%, or 31 individuals, have depolarized their opinions, i.e., they have changed opinions away from their initial opinion. When we combine polarization and moderation with the direction of the opinion change – towards pro or anti positions -, we obtain four groups of “big opinion changers”, as displayed in table 1.

[Table 1 about here]

Graph 1 shows the development of opinions on immigration in the four groups in comparison with “average opinion changers”. While the graph displays clear differences in the magnitude of opinion change between the “big” and “average” changers, we also see that polarization and moderation generally occurs within “ideological camps”: only in group 3 where participants moderated their opinion in pro-immigration directions, opinions shifted over the mid-point of the scale (7) towards the other “ideological camp”.

[Graph 1 about here]

Several comparisons among these four groups will be made. First, we will link opinion change in the four groups to individual-level and group-related variables. Second, we will compare the four groups of “big opinion changers” in the context of polarizing individuals and moderating individuals (comparisons between group 1, group 2, group 3, group 4 and average opinion changers).

### *Operationalization of the predictor variables*

First, we consider education, and knowledge on the issue. With regard to education, there is no standardized definition for low and high education. We have chosen to define elementary, vocational and secondary school as low education and upper college-level, polytechnic and lower and upper university degree as high education. High political interest indicates very or somewhat interested whereas only a little or not at all interested is considered as low political interest. The same logic is used for the variable discuss politics: discuss politics often implies every day or often, whereas sometimes or rarely qualifies as discussing politics rarely. Immigration knowledge is measured through a quiz consisting of 10 questions about immigration in Finland. In the analysis we use the average per cent right answers. In order to capture both initial levels of knowledge and learning, we employ a measure for immigration knowledge before and after discussion.

Next, we focus on empathy, understanding the other side and trust. Empathy and trust are sum variables varying between 0-100 where 0 indicates the lowest level and 100 the highest. Understanding the other side is operationalized via the survey question “someone who does not agree with me on immigration may have good grounds for their opinion”. High levels indicate high understanding whereas low levels indicate low understanding. As with knowledge, we check for the variables before and after discussion.

With regard to ideology we focus, on the one hand, on left-right ideology. The participants were asked to place themselves on a scale, where 0 means the left and 10 means the right. On the other hand, we also consider how the participants feel about immigrants. This is measured in terms of openness and closedness towards immigrants. We use a summary variable, consisting of five questions. All five items load on a factor with an Eigenvalue of 3.121, whereby loadings are high (ranging between .675 and .891). The variable can vary between 0-100 where 0 indicates closedness and 100 indicates maximum openness.

Group context includes variables with regard to the amount of immigrants in the discussion group and participants’ subjective experiences of the discussion. The variable “experienced group pressure” is a summary variable of three questions: a) some participants dominated the discussion too much, b)

other participants interrupted when it was my turn to speak and c) I found it difficult to listen to people who disagree with me. The summary variable can vary between 0-100 and 0 indicates the highest level of experienced group pressure and 100 the absence of group pressure. For the variables high values indicate a good experience whereas low levels indicate a bad experience. The variable “enclaves” indicates how many percent of the individuals was part of a like-minded group compared to a mixed group. For more information about how the variables are coded see the online Appendix.

## Results

We report our results in a simple and easy-to-read way, by comparing the effects of the various individual-level and group-related variables among various groups on the basis of non-parametric tests (see below). A Kruskal-Wallis-Test is used to test whether there are significant differences between the groups for the ordinal variables, while Fisher's Exact Test is used to test the significance level for the nominal variables. We re-check our findings by using more complex statistical models – multinomial probit models -, controlling for covariates simultaneously (see online Appendix). Given the fact that the number of individuals that polarized or moderated is quite small, using a large number of covariates in the regression analyses yields inefficient estimations. Therefore, the multinomial probit models only employ a reduced set of theoretically relevant and significant control variables. However, the results of multinomial probit models largely corroborate the findings of the non-parametric tests (see online Appendix).

As shown in table 2 there are only few differences among polarizing, moderating and “average opinion changers”. Most interestingly, enclave and mixed groups do not drive polarization and moderation. While the amount of polarizing individuals is slightly higher in enclaves compared to mixed groups (64% vs. 53%), this difference is not statistically significant.

[Table 2 about here]

The statistically significant differences among the groups are age, gender and left-right ideology. “Average opinion changers” have the lowest average age, while those who have moderated their opinions are the oldest. In comparison with “average opinion changers”, polarizing tendencies are more common among women. When compared to those participants who moderated their opinions, polarization also seems associated with lower education, even though this result is only found in the regression models (see online Appendix). Ideology matters as well: moderating trends are more strongly associated with left-wing and progressive people. We do not strongly interpret this result, even though a new study on U.S. political blog posts shows that liberal bloggers have a higher level of

cognitive or integrative complexity compared to conservative bloggers, which the authors relate to deliberative quality (Brundigde et al., 2014). Overall, it is surprising that there are so few differences among the groups. Based on previous research on deliberation and group polarization, our expectation was that those moderating their opinions would have higher education, more knowledge, more empathy, and more understanding for different opinions. We also expected that enclaves would fuel polarization. But all that is not the case: various individual-level and group-related factors are no good predictors for polarization and moderation. In the regression models, we have also probed for various interaction effects, such as enclave/mixed groups and education, immigration knowledge and empathy but found none (results upon request).

Next, we focus on different types of polarizing and moderating individuals, namely by drawing comparisons between those who reinforced their negative opinion towards immigration and those who reinforced their positive opinion towards immigration. With five groups in the analysis ten pairwise comparisons have to be made. In general, a Mann-Whitney-Test with Bonferroni correction is employed in such a situation. However, with so many comparisons Bonferroni can be seen as relatively severe. Therefore, we have chosen to use Conover's Post Hoc test that is more permissive and better suited to test significances for more than four groups. Significant differences between any of the groups are shown in the rightmost columns of Table 3.

[Table 3 about here]

Table 3 displays that there are striking similarities between those who polarized opinions toward anti-immigration positions and those who depolarized toward pro-immigration positions (columns 1 and 3) as well between those who polarized opinions toward pro-immigration positions and those who depolarized toward anti-immigration positions (columns 2 and 4). Individuals in columns 1 and 3 have lower levels of education, lower external efficacy, and clearly less positive feelings towards immigrants than individuals in columns 2 and 4 as well as average changers (column 5). They also had less initial immigration knowledge than individuals in columns 2 and 4. In other words, unexpectedly, the same factors seem to predict both polarization and moderation.

What seems puzzling at first glance can be easily solved by taking ideological factors into account. Remind that people in columns 1 and 3 are in the anti-immigration camp at T1, whereas people in columns 2 and 4 are in the pro-immigration camp (see graph 1). Even though there were polarization and moderation trends in both ideological camps, on average the “big opinion changers” still ended up in the same ideological camp after deliberation (with the exception of individuals in column 3 moving to a post-deliberative score of 7.76, even though there is still a significant difference to those in column 4 who depolarized in anti-immigration directions, with a post-deliberative score of 9.25). This interpretation is corroborated by the fact that there are large differences for the variable “feelings about immigrants”, with far less positive feelings in columns 1 and 3. Overall, it is not so surprising that people in the pro-immigration camp have higher education and higher levels of efficacy than people in the anti-immigration camp. The surprising aspect, however, is that polarizing and moderating trends can be observed in both camps. Another important finding is that some differences among the four groups are considerably reduced after the deliberation process: while efficacy levels still differ, knowledge on immigration is almost equalized. There were significant learning effects for all groups (see online Appendix), and especially those with low initial knowledge learned the most. This finding carries significant normative meaning, to which we shall return below.

These ideological underpinnings of polarization and moderation notwithstanding, there is one notable effect with practical ramifications. Those who polarized their opinions in anti-immigrant directions very rarely had an immigrant in their discussion group. Only one of 14 individuals (7%) polarizing their opinions in anti-immigrant directions faced an immigrant in the discussion, whereas those who polarized their opinions in pro-immigration directions, 13 of 22 individuals (59%) had an immigrant present in their discussion group. This finding is in line with the argument that the presence of the *other* – here an immigrant in the discussion group - might have an important effect on how opinions change.

Overall, our results put some question marks on the conventional understandings regarding the normative desirability of moderation and polarization. As mentioned in the theoretical part, the



moderation of opinions is frequently seen as a desirable result of deliberation, while polarization is considered a suspicious outcome. Our findings challenge these widely held interpretations. First, with regard to moderation, the results tend to conform with deliberative expectations, or at least do not contradict desirable deliberative standards for opinion change. Let us start with those participants moderating their views in pro-immigrant directions. Their lower level of education and relatively low initial knowledge on immigration issues raises some questions whether their opinion changes are in line with a deliberative pathway to belief revision. However, they learned quite a bit during discussion even though they still knew least about the immigration issue after deliberation. Moreover, their levels of empathy and understanding were not different from the other groups and with more than 60 points (out of 100) substantively quite high as well; and there were no sign of group pressures. As such, moderation is connected to clearly desirable features. Next, the fact that some well-educated, fairly well-informed, largely empathetic and high-trusting people moderated their opinions toward anti-immigrant positions raises some interesting normative questions. One could say that moderation on such cognitive and ethical grounds is exactly what deliberative theorists vie for. Yet from the perspective of “progressive vanguardism” (see Neblo 2007), this moderation trend is questionable. Individuals with more progressive opinions, as in column 4, should certainly listen to “anti-immigrant” arguments but not necessarily shift their opinions in this direction, since humanitarian positions are expected to have a higher level of universalizability and should therefore have a higher persuasive capacity. As mentioned before, not only is “progressive vanguardism” a contestable position, those moderating their opinions in anti-immigrant directions also ended up in the pro-immigration camp after deliberation (on the 0-14 immigration scale, their post-deliberative opinions were at 9.25). Thus, we would still qualify this moderation trend as largely consistent with deliberative ideals.

When it comes to the polarization of opinions, the picture is not fundamentally different from the one we obtained for moderation. At first glance, those individuals who polarized towards anti-immigration positions might conform to the conventional view. Their level of education and their initial knowledge on immigration issues were lower compared to those individuals who moderated in anti-immigration directions. Yet, they learned quite a bit during discussion and equalized their knowledge compared to the other groups. Moreover, there were neither group pressures nor differing levels of empathy and understanding in comparison with the other groups. Under such conditions, it might be difficult to judge polarization as normatively problematic. Interesting are also those individuals who polarized towards pro-immigration directions. Not only was their level of education and their initial knowledge on immigration issues higher compared to those individuals who polarized in anti-immigration directions (albeit lower than in the group moderating toward anti-immigration positions), they also were more likely to have an immigrant in their small group. Even though they experienced a bit more

group pressure than others, it might be difficult to discard this polarization trend as normatively questionable.

In conclusion, we have both troubling and good news for deliberative theorists. The story behind our results is a bit like the one presented by Sanders (2012) for a transnational deliberative poll (Europolis) – participants change minds after deliberation (and in our case we explicitly focused on those who did the most), but we do not really know why that is the case, even though we have probed for a wide variety of factors including novel measures such as empathy. There is broad agreement in deliberative theory that normatively desirable opinion changes should at least reflect a high epistemic quality and respective capacities of participants, the absence of group pressures, or some ethical aspects such as empathy and understanding. In this regard, our results reveal an intriguing picture: polarizers and moderators were not fundamentally different on epistemic grounds after deliberation, that is, they all learned, and those with low initial knowledge learned even more, group factors barely mattered, and levels of empathy and understanding did not differ among moderators and polarizers either. This is good news for the growing number of deliberative democrats arguing that polarization may reflect preference clarification in that participants better understand what they really want (e.g., Knight and Johnson, 2011). In other words, polarization is not necessarily anti-deliberative.

## **Conclusion**

This paper explored the drivers behind opinion polarization and opinion moderation. Focusing on a citizen deliberative experiment on immigration in Finland, we have analyzed participants who have changed their minds more than average, either in a more extreme or a more moderate direction. The results are quite striking: neither individual-level nor group-related factors are good predictors for the polarization and the moderation of opinions. There are, however, some differences between individuals polarizing in pro- or anti-immigrant directions. But the differences we found are due to basic ideological pre-dispositions. Nonetheless, we found one important factor for opinion polarization in anti-immigrant directions: the absence of immigrants in the small group discussion, which is in accordance with longstanding claims regarding the importance of presence for democratic politics (Philips 1995). The key finding of this study is, however, that polarizers and moderators did not fundamentally differ with regard to epistemic, ethical, and group-related factors (after the deliberative process). As such, our study challenges conventional interpretations of opinion polarization and moderation. While polarization might be problematic from a democratic point of view *at the aggregate level* (since societies and polities tend to fare better when polarization is not extreme), moderation and polarization are not normatively good or bad *per se at the individual level*. As long as these

pathways involve epistemic advancement and ethical aspects, and are not heavily influenced by group dynamics, both polarization and moderation can have deliberative dimensions. This is fully in line with recent advances in philosophy and psychology to understand polarization not as irrational behavior, but – given circumstances and pathways – as an outcome which may conform to normative conceptions of belief revision.

We acknowledge that this study has several limitations. While we have considered a large array of factors that could potentially affect the polarization or moderation of opinions, our list is far from exhaustive. For instance, detailed analyses of the discussions could tell us more about what was really going on in the groups (see Gerber et al. 2014). We have performed such an analysis focusing on equality and rationality in discussion (reference withheld); but due to space considerations we can only provide a summary analysis. The analysis of the group discussions corroborate our findings that polarization is not "irrational" behaviour. We found no differences between the various groups of polarizers, moderators, and "average changers" with regard to speech activity, suggesting that inequalities in presenting arguments was not a driving force behind polarization or moderation. Contrary to conventional expectations, polarizing individuals were not more disrespectful than moderating individuals and "average changers"; interestingly, moderating individuals were even slightly less respectful than polarizing ones. Moreover, polarizing individuals even presented the largest proportion of rational arguments, even though this pattern only occurs for one of the polarizing groups.

Overall, the fact that frequently mentioned sources for opinion polarization or moderation– such as group dynamics or empathy and understanding – did not matter in this experiment has major implications for future research. Indeed, people do change minds in deliberative events (and even massively), and we need to devise a more dedicated research program to understand why that is the case. Otherwise, the big normative claims surrounding deliberation, such as well-considered judgments and decision legitimacy, eventually rest on shaky foundations.

## References

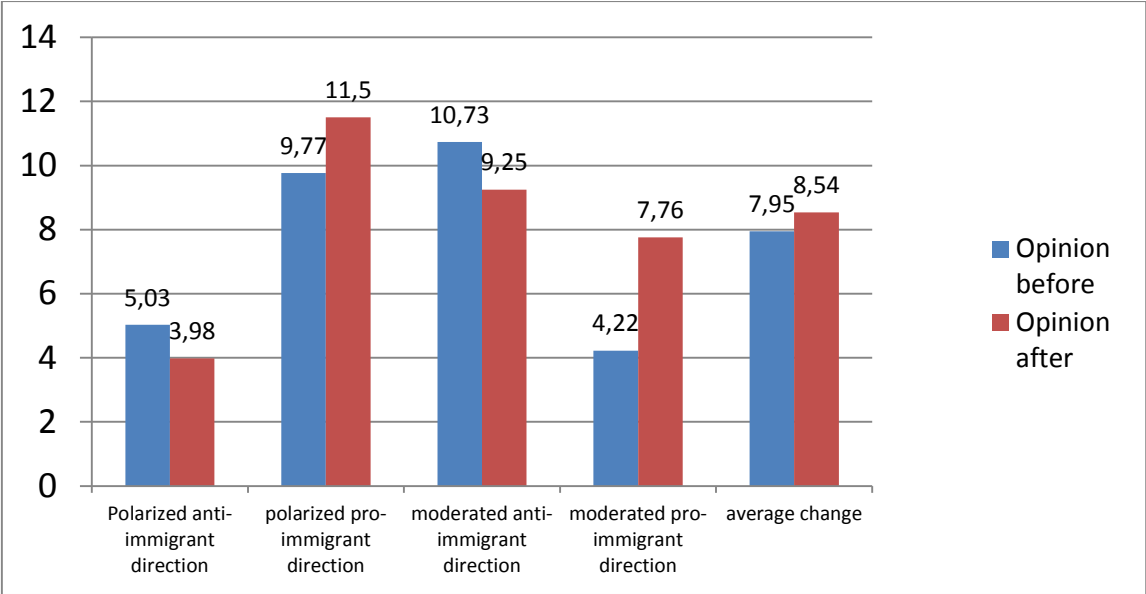
- Ackerman, B. & Fishkin, J. S. (2004) *Deliberation Day*, New Haven & London: Yale University Press.
- Andersen, V. & Hansen, K. (2007) "How deliberation makes better citizens: The Danish Deliberative Poll on the euro", *European Journal of Political Research*, vol. 46, no. 4, pp. 531-556.
- Baccaro, L., Bächtiger, A. & Deville, M. (2014) "Small Differences That Matter: The Impact of Discussion Modalities on Deliberative Outcomes", *British Journal of Political Science*, forthcoming.
- Barabas, J. (2004) "How Deliberation Affects Policy Opinions", *American Political Science Review* vol. 98, November, pp. 687-701.
- Blais, A., Carty, K. R. & Fournier, P. (2008) "Do citizens' assemblies make reasoned choices?", in Warren, M. E. & Pearse, H. (ed) *Designing Deliberative Democracy. The British Columbia Citizens' Assembly*, New York: Cambridge University Press.
- Brundidge, J. A., Scott, R., Choi, S. & Muddiman, A. (2014) "The "Deliberative Digital Divide:" Opinion Leadership and Integrative Complexity in the U.S. Political Blogosphere", *Political Psychology*, forthcoming.
- Bruneau, E. G., Dufour, N. & Saxe, R. (2012) "Social Cognition in Members of Conflict Groups: Behavioural and Neural Responses in Arabs, Israelis and South Americans to Each Other's Misfortunes", *Philosophical Transactions of the Royal Society B* vol. 367, no. 1589, pp. 717-730.
- Cikara, M., Bruneau, E. G. & Saxe R. (2011) "Us and Them: Intergroup Failures of Empathy", *Current Directions in Psychological Science* vol. 20, no. 3, pp. 149-153.
- Davis, M. H. (1980) "A Multidimensional Approach to Individual Differences in Empathy", *JSAS Catalog of Selected Documents in Psychology* 10.
- Chambers, S. (2003) "Deliberative Democracy Theory", *Annual Review of Political Science*, vol. 6, pp. 307-326
- Denver, D., Hands, G. & Jones, B. (1995) "Fishkin and the deliberative opinion poll: Lessons from a study of the Granada 500 television program", *Political Communication*, vol. 12, no. 2, pp. 147-156.
- Dutwin, D. (2003) "The Character of Deliberation: Equality, Argument, and the Formation of Public Opinion", *International Journal of Public Opinion Research*, vol. 15, no. 3, pp. 239-264.
- Farrar, C., Green, D.P., Green, J.E., Nickerson, D.W. & Shewfelt, S. (2009) "Does Discussion Group Composition Affect Policy Preferences? Results from Three Randomized Experiments", *Political Psychology* vol. 30, no. 4, pp. 615-647.
- Farrar, C., Fishkin, J. S., Green, D. P., List, C., Luskin, R. C. & Paluck, E. (2010) "Disaggregating Deliberation's Effects: An Experiment within a Deliberative Poll", *British Journal of Political Science* vol. 40, pp. 333-347.
- Fishkin, J. S. (1995) *The Voice of the People*, New Haven and London: Yale University Press.
- Fishkin, J.S. & Luskin, R.C. (1999) "Bringing deliberation to the democratic dialogue", in McCombs, M. & Reynolds, A. (ed) *The Poll With A Human Face: The National Issues Convention Experiment in Political Communication*, Mahwah, NJ: Routledge.
- Fishkin, J.S. (2009) *When the People Speak. Deliberative Democracy & Public Consultation*, Oxford: Oxford University Press.
- Forst, R. (2001) "The Rule of Reasons. Three Models of Deliberative democracy", *Ratio Juris*, vol. 14, no 4, pp. 345-378.
- French, D. & Laver, M. (2009) "Participation, Bias, Durable Opinion Shifts and Sabotage through Withdrawal in Citizens' Juries", *Political Studies*, vol. 57, pp. 422-450.
- Gerber, M., Bächtiger, A., Fiket, I., Steenbergen, M. & Steiner, J. (2014) "Deliberative and non-deliberative persuasion: Mechanisms of opinion formation in EuroPolis", *European Union Politics*, vol. 15, no. 3, September, pp. 410-429.
- Hall, T. E., Wilson, P. and Newman, J. (2011) 'Evaluating the Short- and Long-term Effects of a Modified Deliberative Poll on Idahoans' Attitudes and Civic Engagement Related to Energy Options', *Journal of Public Deliberation*, vol. 7, no. 1, pp. 1-30.

- Hoffman, M.L. (2000) "Empathy and Moral Development. Implications for Caring and Justice", Cambridge: Cambridge University Press.
- Hogg, M. A. (1993) "Group Cohesiveness: A Critical Review and Some New Directions", *European Review of Social Psychology*, vol. 4, no. 1, pp. 85-111.
- Isenberg, D.J. (1986) "Group Polarization: A Critical Review and Meta-Analysis", *Journal of Personality and Social Psychology* vol. 50, no. 6, pp. 1141-1151.
- Jones, D.A. (2013) "The Polarizing Effect of a Partisan Workplace", *PS: Political Science & Politics* vol. 46, no. 1, pp. 67-73.
- Karpowitz, C. F., Raphael, C. & Hammond IV, A. S. (2009) "Deliberative Democracy and Inequality: Two Cheers for Enclave Deliberation among the Disempowered", *Politics & Society* vol. 37, pp. 576-615.
- Knight, J. & Johnson, J. (2011) *The Priority of Democracy: Political Consequences of Pragmatism*, Princeton: Princeton University Press.
- Kriesi, H., Grande E., Lachat, R., Dolezal, M., Bornschieer, S. and Frey, T. (2006) 'Globalization and the transformation of the national political space: six European countries compared', *European Journal of Political Research*, vol. 45, no. 6, pp. 921-957.
- Kunda, Z. (1990) "The Case for Motivated Reasoning", *Psychological Bulletin* vol. 108, no. 3, pp. 480-498.
- Lang, A. (2007) "But is It for Real? The British Columbia Citizens' Assembly as a Model of State-Sponsored Citizen Empowerment", *Politics & Society*, vol. 35, pp. 35-69.
- Luskin, R. C., Fishkin, J. S. & Jowell, R. (2002) "Considered Opinions: Deliberative Polling in Britain", *British Journal of Political Science*, vol. 32, pp. 455-487.
- Mansbridge, J., Bohman, J., Chambers, S., Christiano, T., Fung, A., Parkinson, J., Thompson, D. & Warren, M. (2012) "A systemic approach to deliberative democracy", in Parkinson, J. & Mansbridge, J. (ed) *Deliberative Systems*, Cambridge University Press.
- Mercier, H., & Landemore, H. (2012) "Reasoning is for Arguing: Understanding the Successes and Failures of Deliberation", *Political Psychology* vol. 33, no. 2, pp. 243-258.
- Mendelberg, T. (2002) "The Deliberative Citizen: Theory and Evidence", *Political Decision Making, Deliberation and Participation*, vol. 6, pp. 151-193.
- Merkle, D. (1996) "The National Issues Convention Deliberative Poll", *Public Opinion Quarterly*, vol. 60, pp. 588-619.
- Morrell, M. (2010) *Empathy and Democracy. Feeling, Thinking and Deliberation*, University Park: Penn State University Press.
- Mutz, D. (2006) *Hearing the Other Side. Deliberative versus Participatory Democracy*, New York: Cambridge University Press.
- Neblo, M. (2007) "Family disputes: Diversity in defining and measuring deliberation", *Swiss Political Science review*, vol. 13, no. 4, pp. 527-557.
- Philips, A. (1995) *The Politics of Presence*, Oxford: Clarendon Press.
- Preston, S, D. and de Waal F. B. M. (2002) "Empathy: Its Ultimate and Proximate Bases", *Behavioral and Brain Sciences* vol. 25, no. 1, pp. 1-72.
- Reykowski, J. (2006) "Deliberative democracy and "Human Nature": An Empirical Approach", *Political Psychology*, vol. 27, pp. 323-346.
- Rosenberg, S. (2005) "The empirical Study of Deliberative Democracy: Setting a Research Agenda", *Acta Politica*, vol. 40, pp. 212-224.
- Rosenberg, S. (2007) "Rethinking Democratic Deliberation: the Limits and Potential of Citizen Participation", *Polity*, vol. 39, pp. 335-360.
- Sanders, D. (2012) "The effects of deliberative polling in an EU-wide experiment: Five mechanisms in search of an explanation", *British Journal of Political Science*, vol. 42, no. 3, pp. 617-640.
- Schweigert, F. J. (2010) "Strengthening Citizenship Through Deliberative Polling", *Journal of Community Practice*, vol. 18, pp. 19-39.
- Setälä, M., Grönlund, K. & Herne, K. (2010) "Citizen Deliberation on Nuclear Power: A Comparison of Two Decision-Making Methods", *Political Studies*, vol. 58, no. 4, Oktober, pp. 688-714.

- Sunstein, C. (2002) "The Law of Group Polarization", *The Journal of Political Psychology*, vol. 10, pp. 175-195.
- Sunstein, C. (2006) *Infotopia: How many Minds Produce Knowledge*, New York: Oxford University Press.
- Sunstein, C. (2009) *Going to Extremes. How Like Minds Unite and Divide*, New York: Oxford University Press.
- Uslaner, E. (2001) "Producing and Consuming Trust", *Political Science Quarterly*, vol. 115, pp. 569-590.
- Walter, Henrik (2012) "Social Cognitive Neuroscience of Empathy: Concepts, Circuits, and Genes", *Emotion Review*, vol. 4, no. 1, pp. 9-17.
- Verba, S., Schlozman, K.L. & Brady, H.E. (1995) *Voice and Equality – Civic Voluntarism and American Politics*, Cambridge: Harvard University Press.
- Vinokur & Burnstein (1978) "Moderation of Attitudes in groups", *Journal of Personality and Social Psychology*, vol. 26, pp. 872-885.
- Wojcieszak, M. (2011) "Deliberation and Attitude Polarization", *Journal of Communication*, vol. 61, pp. 596-617.

**Table 1. Four groups of “big opinion changers”**

Initial opinion and direction of change	n	Direction of opinion change
Group 1: Positive initial opinion toward immigration and positive change	22	Polarization
Group 2: Negative initial opinion toward immigration and negative change	14	Polarization
Group 3: Positive initial opinion toward immigration and negative change	17	Moderation
Group 4: Negative initial opinion toward immigration and positive change	13	Moderation
Total	66 (of 207)	



**Graph 1. Opinion change for the “big opinion changers” and average changers**

**Table 2. Comparisons between polarizers, moderators and average opinion changers**

	Polarizing individuals (N=36)	Moderating individuals (N=30)	Average changers (N=141)	pol-mod	pol-average	mod-average
<b>INDIVIDUAL-LEVEL FACTORS</b>						
<b>Average age**</b>	55	61	50	†		**
<b>Men (%)<sup>1†</sup></b>	33	40	53		*	
High educated (%) <sup>1</sup>	33	45	43			
Education (1-8)	4.4	4.7	4.9			
Immigration knowledge -pre	44	39	43			
Immigration knowledge -post <sup>†</sup>	61	59	65			†
Empathy -pre	68	65	68			
Empathy -post	70	66	68			
Social trust -pre	62	64	65			
Social trust-post	59	65	62			
Feelings about immigrants	51	53	50			
<b>Ideology (0-10)*</b>	4.3	5.9	4.8	**		*
<b>GROUP CONTEXT</b>						
Like-minded treatment (%) <sup>1</sup>	64	53	58			
Immigrant in group (%) <sup>1</sup>	39	47	43			
Experienced group pressure	23	31	28			

Notes: <sup>1</sup>Significance test: Fisher's Exact Test, other variables: Kruskal Wallis Test: †p<0.10, \*p<.05, \*\*p<.01, \*\*\*p<.001. Columns to the right show significant pairwise differences (Conover's Post Hoc Test). We also performed analysis distinguishing between cognitive and emotional empathy but results did not change.



**Table 3. Comparisons of different groups of polarizers and moderators**

	(1) Polarization toward anti-immigration positions (n=14)	(2) Polarization toward pro-immigration positions (n=22)	(3) Moderation toward pro-immigration positions (n=13)	(4) Moderation toward anti-immigration positions (n=17)	(5) Average changers	1-2	2-3	2-4	1-3	1-4	3-4	1-5	2-5	3-5	4-5
INDIVIDUAL-LEVEL FACTORS															
Average age*	56	55	63	59	50		†							**	*
Men (%) <sup>1</sup>	29	36	46	35	53										
High educated (%) <sup>1†</sup>	21	41	17	65	43				*	*					
Education (1-8)**	4.0	4.6	3.4	5.7	4.9		†	†	**	**	†			*	†
Immigration knowledge-pre †	39	48	35	42	43	†	**						†	†	
Immigration knowledge – post*	59	62	52	64	65					†	†			**	
Empathy -pre	63	72	64	66	68										
Empathy -post	66	73	65	67	68										
Social trust –pre**	50	69	55	70	65	**	*		**	*	*			†	
Social trust –post**	47	66	58	71	62	**			***	*	**				*
Feelings about immigrants***	30	65	36	66	50	***	***		***	***	***	***	***	**	***
Ideology (0-10)*	4.5	4.2	6.5	5.5	4.8		**	†	*					*	
GROUP CONTEXT															
Like-minded treatment (%) <sup>1</sup>	43	77	39	65	57										
Immigrant in group (%) <sup>1*</sup>	7	59	46	47	43	**			†	†		*			
Experienced group pressure*	32	17	33	30	28	**	*	*					**		

Notes: <sup>1</sup>Significance test: Fisher’s Exact Test, other variables: Kruskal Wallis Test: †p<0.10, \*p<.05, \*\*p<.01, \*\*\*p<.001. Columns to the right show significant pairwise differences (Conover’s Post Hoc Test)