Political Studies Association

politics

POLITICS: 2014 VOL 34(2), 101–115

doi: 10.1111/1467-9256.12043

Research Article

Does Inter-group Deliberation Foster Inter-group Appreciation? Evidence from Two Experiments in Belgium

Didier Caluwaerts

Vrije Universiteit Brussel and Harvard University

Min Reuchamps

Université Catholique de Louvain

Deliberative democrats assume that political deliberation is capable of transforming citizens' opinions and attitudes. This article takes this assumption as a starting point and tries to test it empirically by determining whether deliberation in an inter-group setting induces more positive out-group attitudes. Based on data from two deliberative experiments in Belgium, we argue that the overall effect of deliberative quality on attitude change is limited. The most important determinant of changes in out-group attitudes is the group composition. Citizens who are confronted with the out-group are more likely to hold more positive out-group attitudes afterwards.

Keywords: deliberative democracy; Belgium; inter-group relations; contact hypothesis

When political issues have long generated deep divisions, or when public debates touch upon fundamental issues of identities and interests, the willingness to engage in a constructive dialogue may prove to be rather limited. This is especially the case when public deliberation takes place in an inter-group setting: when citizens are confronted with their out-group, they activate stereotypes and may be reluctant to review previously held opinions and feelings towards that out-group (Hewstone, Rubin and Willis, 2002; Mendelberg, 2002).

But a long tradition of research, building on Gordon Allport's (1954) contact hypothesis, suggests that communication between conflicting groups might lead to more tolerant and cosmopolitan decision outcomes (Gastil, Bacci and Dollinger, 2010) and positive attitude changes (Chambers, 2003; Mutz, 2006), as well as more integrative solutions (Maoz and Ellis, 2008). What remains to be disentangled, however, is whether these effects are brought about by the qualities of the interactions or by the mere fact of facing the out-group.

This article therefore sets out to determine whether inter-group deliberation fosters inter-group appreciation. More specifically, our research question is twofold. On the one hand, we want to know whether a high quality of democratic deliberation can lead to a more positive regard for members of the out-group. On the other hand, we ask whether these transformative effects also take place when deliberants are confronted with the out-group. Our findings suggest that the intrinsic qualities of the deliberative process are only weakly

linked to changes in out-group attitudes, and that the main variable explaining shifts towards more positive regard for the out-group is actually group composition. Participants who are confronted with the other side move to more positive views of the out-group regardless of the quality of the deliberation.

We start this article with a discussion of the relationship between deliberation, inter-group contact and attitude change. Afterwards, we present the background and methodology of two deliberative mini-publics organised in Belgium. Finally, we turn to our results and draw conclusions on the potential effects of deliberative democracy in conflict-ridden societies.

1 Inter-group deliberation and attitude transformation

The transformative effects of citizen deliberation have long been assumed in deliberative theory (Dryzek, 2000), and many empirical studies have hitherto successfully demonstrated that citizens are not deterred from changing their opinions during deliberation (Chambers, 2003; Gastil, Bacci and Dollinger, 2010). They are capable of opening up their minds, and often go through a deliberative learning curve. A significant contribution on the dynamics of social learning through deliberation was made by the research on deliberative polls (Fishkin, 1995; Luskin et al., 2012). Deliberative polls start from the idea that regular opinion polls rarely reflect the informed preferences of the public, and that deliberation can enhance the consideredness of public opinions. Through measuring citizens' opinions and attitudes in pre-test and post-test questionnaires, deliberative polls consistently show that citizens hold different opinions after having the opportunity to talk to each other.

Moreover, research has also found some directional effects. Deliberators not only change their minds, they generally also become more tolerant of different opinions and they tend to find common ground where there used to be conflict. For instance, John Gastil, Chiara Bacci and Michael Dollinger (2010) argue that attitude changes are rarely neutral, and that deliberation fosters more cosmopolitan and collectivist value orientations. Participants incorporate others' views and set aside particularistic or nationalistic rationales in favour of feelings of commonality and mutual trust (Gastil, Bacci and Dollinger, 2010, p. 5; Luskin et al., 2012). These findings lead to the following hypothesis:

H1. A high quality of deliberation will lead to more positive out-group attitudes.

However, the quality of deliberation may have little effect when there is strong antagonism in the group. The question, then, is whether deliberants will be willing to change their minds, even when they are confronted with conflicting views. After all, when deliberation takes place between diametrically opposing groups in society, the dynamics of deliberation are bound to be different.

Thus, the issue is whether the first hypothesis holds true when citizens are confronted with their out-group. Some scholars argue that inter-group deliberation will reinforce rather than reduce conflicts. For instance, Magdalena Wojcieszak (2012) finds that the presence of 'strong attitudes' may matter much more than the quality of deliberation, and that deep conflicts can undermine any transformative effect, even if the quality of deliberation is high. Michael Morrell (2010) adds to this that preference transformation requires the capacity to move oneself into the position of others, and to be sensitive to others' feelings. It requires empathy, which is not so easily achieved in an inter-group setting.

Others argue that inter-group deliberation can foster the reduction of inter-group conflicts. This argument finds support in the contact hypothesis (Allport, 1954; Pettigrew and Tropp, 2006), which acknowledges that stereotypes and prejudices form the basis of inter-group conflict, but also states that contact between competing groups reduces inter-group tensions (Dovidio, Gaertner and Kawakami, 2003). Putting groups together is thus claimed to foster inter-group appreciation, an argument that resonates in communication studies, where it is found that discussing political issues with diverse others leads to greater trust and a better regard for out-groups (Cappella, Price and Nir, 2002; Mutz, 2006).

Interesting on this point are examples of inter-group deliberations between Israeli Jews and Palestinians, and between Protestants and Catholics in Northern Ireland. In the first example, Donald Ellis and Ifat Maoz (2002; 2007) have shown that inter-group communication increases the propensity for peaceful conflict resolution, due to the mediating effect of out-group trust, bearing in mind that the majority/minority power relations were better predictors of the inter-group dynamics than their respective cultural communication codes. In the latter case, a deliberative poll in Northern Ireland showed that the deliberators' attitudes towards the other religious group shifted, and that they became more convinced of the trustworthiness of the out-group after talking to them, despite the long history of hostilities between Protestants and Catholics (Luskin et al., 2012). These findings are also reflected in the recent Europolis deliberative poll, which gathered European citizens to discuss migration policies. It showed that citizens with highly diverse opinions are able to come to some kind of reasoned understanding (Fiket, Olsen and Trenz, 2011).

Deliberation thus seems possible in high-conflict settings, and inter-group deliberation may foster inter-group appreciation, but only when the majority/minority dynamic is controlled for with an equal composition (Meyers, Brashers and Hanner, 2000). Combined with the insights from the contact hypothesis, these previous studies allow us to hypothesise the following:

H2. When deliberation takes place in a divided setting, a mixed but equal group composition will foster a change towards more positive out-group feelings.

2 Evidence from two deliberative experiments

Testing these hypotheses requires the observation of communicative interactions between citizens from opposing groups. Due to their highly segmented nature, however, divided societies are characterised by only a limited number of contacts between groups (O'Flynn, 2006). This inherent limitation explains why we rely on data from two deliberative experiments organised in Belgium in 2010 and 2011.

Belgium is an exceptional case for studying the effects of inter-group deliberation. Despite the absence of violent conflict, it is one of the most segmented countries imaginable: it has two completely separate public and political spheres, and the possibilities and incentives for Belgian citizens to inform themselves about the viewpoints of the other linguistic group are limited (Caluwaerts, 2012; Reuchamps, 2013; Reuchamps, Kavadias and Deschouwer, 2013). As such, the threshold for mutual contact between Dutch and French speakers is high. This was even a deliberate strategy on the part of Belgian political elites because it was feared that inter-group contact would only increase mutual hostilities (Bogaards, 1998). Yet any nation-wide political decision has to be made jointly by – the representatives of – both groups.

Moreover, at the moment when these two experiments were organised, Belgium was stuck in a long political crisis and tensions between the two linguistic groups were at a historic high (Deschouwer and Reuchamps, 2013). It is in this political context that the first experiment gathered ordinary Belgian citizens from both sides of the linguistic cleavage in nine deliberative groups. Three of these groups were homogeneously Dutch speaking, three others were homogeneously French speaking, and the final three were composed of both French and Dutch speakers. Theoretically, all of these groups consisted of 10 persons, ensuring a gender balance, and a balance between the linguistic groups in the divided groups.[1] In practice, however, 83 out of 90 invitees eventually attended.

These participants were recruited by administering a pre-test survey to a random but disproportionately stratified sample ($n = 2{,}024$) of Belgian citizens.[2] Due to financial restrictions, however, the sample was drawn from an existing panel with over 110,000 individuals rather than official census lists, but the panel closely resembled the socio-demographic composition of the Belgian population.[3]

The random selection was not paralleled by a fully random assignment of the participants to the groups. Given the small group size, random assignment has a high chance of generating groups that are internally homogeneous (Caluwaerts and Ugarriza, 2012). This is not desirable from a theoretical point of view, or in light of our research question. The participants were therefore assigned to the groups based on a randomised block sample defining the blocks based on gender, age and language. *Ex post* sample checks revealed that the composition of each group was comparable (Caluwaerts, 2012).

The question we presented to the participants was as simple as it was controversial: 'how do you see the future of Belgium?' At a moment when the negotiations on the state reform had stalled, and early elections were in sight, this deliberative experiment dealt with issues that went to the heart of the Belgian political deadlock. The participants gathered for half a day to discuss this hot topic. Even though the participants were informed that their results would be submitted to a parliamentary committee and to the media, there were only limited consequences for actual politics, which made it a relatively 'cold' deliberative setting (Fung, 2003). Moreover, they did not receive any briefing materials, and the moderator merely introduced the experiment, and opened and closed the discussion following the same format for each of the nine groups, to ensure a high internal validity.

The second mini-public took place a year later, during the political crisis after the 2010 elections, when Flemish nationalists won the elections and the formation of a new government took 541 days. This time, 704 ordinary Belgian citizens participated but the methodology of the G1000 Citizens' Summit, as it was called, was largely comparable to that of the first study. As in the first experiment, the participants were seated at either a linguistically homogeneous table (30 Dutch- and 19 French-speaking tables) or at one of the 32 bilingual tables. Each of these tables consisted of between seven and 10 persons, ensuring a gender balance, and a balance between the linguistic groups at the divided tables.[4]

An independent recruitment agency was asked to contact participants through Random Digit Dialing.[5] To guarantee the quality of the sample, we checked for certain predefined population quotas, such as gender, age and language. The participants were subsequently assigned to the groups using a block randomisation that took the same population quota into consideration. Because we could not offer any financial incentives, there was a – not unusual – dropout rate of about 30 per cent among the people who had previously confirmed their participation, with

Table 1: Comparison of socio-demographics

		Participants Exp. 1	Participants Exp. 2	All Belgians
Sex	Man	53.6%	48.5%	48.4%
	Woman	46.4%	51.8%	51.6%
Age	≤29	26.2%	17.0%	18.8%
	30–49	51.2%	45.0%	35.4%
	50≤	22.6%	38.0%	45.8%
Educ.	Lower	3.6%	7.7%	33.8%
	Secondary	28.6%	22.6%	37.7%
	Higher	67.9%	69.7%	28.5%

704 final members. This final group of participants consisted of 61 per cent Dutch speakers and 39 per cent French speakers. It was diverse even if not perfectly representative in terms of socio-demographics (Table 1).

While the methodology of the two experiments was similar, the topics under discussion in Experiment 2 were broader than the future of Belgium per se. All of the participants discussed the following three topics: social security, distribution of welfare and immigration. These were reached through an open agenda-setting process in which all Belgian citizens were invited to propose issues (Caluwaerts and Reuchamps, 2012; G1000, 2012). Each topic was introduced by two experts and then discussed at each table with the help of a trained moderator. The deliberations for this second experiment took one day.

As in the first study, the purpose of the deliberation was not to reach a consensus. Rather, the participants were asked to formulate a number of policy proposals that could be presented to parliament and the media. Besides their comparative size, the main difference between the two experiments is that the second took place in a much 'hotter' deliberative setting (Fung, 2003). After all, there was much more media coverage and politicians were openly polarised on the G1000.

3 Key variables

In order to answer our research questions, we needed to measure the feelings the participants had towards the out-group. For both experiments, we therefore included the following question in the pre-test and post-test questionnaires: 'Could you indicate, on a scale ranging from 0 to 10 how positive or negative you feel about [the other linguistic group]?' A score of '0' indicated negative feelings, whereas '10' indicated a positive attitude. To determine the opinion change, we subtracted the pre-test score from the post-test score. A positive result indicated becoming more positive, whereas a negative score meant that the participant became more negative.

The deliberative quality was measured using the Discourse Quality Index (DQI) (Steenbergen et al., 2003). This DQI measures interruptions, respect for counter-arguments, the level of justification and constructive politics; these all capture an essential dimension of deliberation.

Table 2: Intercoder reliability tests[a]

	Experiment 1	
	RCA	**Cohen's κ**
Interruption	1	
Respectful language	1	
Respectful listening	0.938	0.646 (0.144)**
Respect for counter-arguments	0.864	0.781 (0.060)**
Level of justification	0.776	0.716 (0.058)**
Abstract principles	0.975	0.787 (0.146)**
Constructive politics	0.902	0.796 (0.068)**

Notes: [a]The RCA indicates the number of identical codes given by the coders, with 1 indicating a perfect overlap. Cohen's kappa additionally takes into account the agreement that would have occurred by chance.
*n = 81; ** P = < 0.01.*

However, because the initial DQI was designed to analyse parliamentary discourse, we added the use of respectful language and respectful listening,[6] two crucial dimensions of citizen deliberation.

For the first experiment, we coded 1,664 speech acts on all of these dimensions, and 81 were re-coded by a trained second coder. This selection of speech acts was purposefully sampled, in order to contain speech acts from both French and Dutch speakers, as well as short and long speech acts. Table 2 reports that the ratio of coding agreement (RCA) (Holsti, 1969) and Cohen's kappa (Cohen, 1960) were high, indicating strong inter-coder reliability.

Because the second experiment was much larger, we randomly selected six tables with varying group compositions and coded a total of 5,152 speech acts. However, because the debates were more structured, the DQI dimensions showed little variation. The active moderation largely appears to have levelled out any differences in DQI scores between the participants, and this inevitably means that we cannot use measurements of the effective quality of deliberation for the second experiment. We will therefore have to rely only on the perception of deliberation, as we will see below.

In order to create the index out of the individual items in the first experiment, we used a principal component analysis. Table 3 reports that five out of seven DQI items load well on one factor, and the analysis of the Kaiser-Meyer Olkin measure indicated that the items were factorable (KMO = 0.753). Respect for counter-arguments, respectful listening and the level of justification are the three strongest items, but they are completed by the use of respectful language and constructive politics, which have slightly lower component loadings. All of these items refer to a way of presenting one's arguments, and defending one's position. Given the positive results of the factor analysis, we created an additive scale excluding interruption and abstract principles. The scale has a good internal consistency and reliability.

In addition to DQI scores based on the transcripts, we also measured the *perceived* quality of deliberation. According to Laura Black et al. (2011) and Jürg Steiner (2012), the perception

Table 3: Component loadings of the Discourse Quality Index

	Component loading
Interruption	0.309
Respectful language	0.526
Respectful listening	0.760
Respect counter-arguments	0.831
Level of justification	0.709
Abstract principles	0.323
Constructive politics	0.608
Eigenvalue	3.618
Percentage of total variance	43.394
Cronbach's alpha (excluding interruption and abstract principles)	0.739

of deliberation might actually be more important than the actual quality of the deliberation. After all, it might well be that a discussion is high on deliberative quality, but if the participants do not perceive it as such, talk might have little effect. The perception of deliberation could thus be a crucial factor in understanding inter-group deliberation.

To measure the participants' perception of the quality of the discussion, we put nine items in the post-test questionnaire, each of them encompassing a theoretical dimension of deliberation, and ran a principal component analysis. The results of the Varimax rotated solutions in Table 4 report that we can distinguish two factors in both experiments, on which the perceived DQI items load well. The first factor contains items referring to the way in which the participants perceive their own deliberative behaviour and efforts. The second factor (PDQI 2) gives an appreciation of the deliberativeness of others in the group. It captures whether others were respectful, truthful or open to being convinced. Analyses of the Kaiser-Meyer Olkin measures showed that the items were factorable, and the measures were reliable in both experiments. As such, additive scales ranging from 0 to 10 were created.

4 Findings

In Experiment 1, before trying to explain attitude change, we should know whether any variation occurs in the participants' attitudes before and after deliberation. One way of making this determination is by looking at the raw changes in mean scores. Table 5 shows that participants held a significantly more positive out-group attitude after deliberation. However, mean group changes can obscure levels of individual change. We therefore also looked at the percentages of the participants changing their feelings towards the out-group. Of all participants 31.3 per cent did not change their opinions about the out-group from the pre-test to the post-test. Nonetheless, more than half of the participants in the experiment held a more positive view of the out-group than before, whereas only 13.3 per cent of them became more negative. Although it is generally assumed that deliberation moves citizens in more pro-social directions, it is still remarkable that the participants shifted so massively in favour of the out-group.

Even though the results in Table 5 show a clear shift in out-group attitudes, we still have to test whether this is due to deliberation. We hypothesised that a high deliberative quality

Table 4: Varimax rotated component loadings of the perception of deliberation

	Experiment 1		Experiment 2	
	Loading component 'PDQI 1'	Loading component 'PDQI 2'	Loading component 'PDQI 1'	Loading component 'PDQI 2'
I did not bring up some of my ideas and viewpoints out of fear of being ridiculed (coded inversely).	0.735		0.800	
I feel that I had to hide my true beliefs (coded inversely).	0.745		0.793	
I had ample opportunities to express my opinion during the discussion.	0.663		0.588	
No matter how hard I tried, the other participants seemed unwilling to listen to what I had to say (coded inversely).	0.785		0.572	
Overall, people were treated with respect during the discussion.		0.462		0.678
Most participants genuinely cared about the common good of all.		0.621		0.639
In general, I understood the arguments of the other participants.		0.595		0.581
In general, I think good arguments were formulated during the discussion.		0.755		0.650
Overall, I feel that people expressed what was truly on their minds.		0.760		0.589
Eigenvalues	3.570	1.453	3.401	1.306
Percentage of total variance	40.559	16.146	39.452	14.395
Cronbach's alpha	0.701	0.641	0.701	0.648

would foster a more positive out-group attitude (H1), and that these effects would hold in an inter-group context (H2). The central variables in this analysis are thus the quality of the deliberation and the group composition.

As control variables, we included gender, age, education and language. Gender was controlled for because there is an ongoing debate about the influence of gender on deliberation: Christopher Karpowitz, Tali Mendelberg and Lee Shaker (2012) argue that deliberation is stratified along the lines of gender, whereas others disagree (Hickerson and Gastil, 2008). The speakers' educational attainment was also included because the higher educated are generally more politically knowledgeable and skilled. This means that their opinions and attitudes may

Table 5: Changes in out-group feelings (individual level)

	Mean pre-test (T1)	Mean post-test (T2)	Mean change (sign. paired samples t-test)	Percentage unchanged	Percentage more positive	Percentage more negative
In general, how positive or negative are you with regard to [the other linguistic group]?	6.61	7.56	0.95***	31.3% (*n* = 26)	55.4% (*n* = 46)	13.3% (*n* = 11)

*Notes: n = 83; *** P < 0.001.*

be relatively steady (Hooghe, 1999), and that they may also be more tolerant of those with whom they disagree (Delli Carpini and Keeter, 1996). We therefore added a dummy variable, which distinguished between those who are higher educated (code '1') and those who received secondary education at most (code '0').[7] The age of the speaker was also considered relevant because older generations experienced nationalist tensions first-hand, whereas the younger generations grew up in times of relative peace. Finally, because the Belgian conflict is a linguistic one, we also included the speaker's native language as a control.

In order to test our hypotheses, we ran a regression analysis predicting the change in out-group feelings (Table 6). Because deliberators are always nested in groups, the data might have been clustered, which is why we calculated the robust standard errors.[8] Surprisingly, the quality of the deliberation (whether effective or perceived) is not significantly related to attitude change: better deliberation does not lead to a more magnanimous attitude towards the out-group. Instead, the crucial factor in explaining changes in out-group feelings is group composition. This is an important finding: facing the out-group leads to a much more favourable out-group attitude, which seems to confirm our second hypothesis. It also lends support to Allport's (1954) contact hypothesis, with the following nuance: even relatively short and unstructured interactions between members of opposing groups can trigger a more positive attitude.

Even though this first experiment had only a small sample, we also interacted the group composition with the DQI in an ANCOVA (not shown here). The results showed that the interaction was not significant, and that the effect of deliberative quality does not differ across the various group settings. The effect of the group composition, however, remained significant at the 0.10 level, which confirms our previous finding that contact with the out-group leads to more positive feelings, regardless of the (perceived) quality of deliberation.

Table 6: OLS regression predicting the change in out-group feelings (T2 – T1) with clustered robust standard errors

	B	S.E	Sign.
Constant	–0.283	2.011	0.894
DQI score	0.146	0.212	0.497
PDQI 1	–0.125	0.148	0.393
PDQI 2	–0.175	0.209	0.379
Divided group	0.821	0.397	0.032
Gender	0.096	0.392	0.817
Age	0.026	0.014	0.072
Higher education	–0.346	0.562	0.539
French speaker	0.446	0.331	0.180
Adjusted R^2		17.9%	

Note: n = 83.

Table 7: Changes in out-group feelings (individual level)

	Mean pre-test	Mean post-test	Mean change (sign. paired samples t-test)	Percentage unchanged position	Percentage becoming more positive	Percentage becoming more negative
In general, how positive or negative are you with regard to [the other linguistic group]?	7.26	7.21	–0.05 (n.s.)	37.7% (*n* = 159)	31.5% (*n* = 133)	30.6% (*n* = 129)

Note: n = 478.

As in the first experiment, in Experiment 2 we first have to determine whether actual shifts in attitudes have taken place. As Table 7 shows, however, there have been no mean changes. The mean pre-test and post-test scores were 7.26 and 7.21, respectively; however, this lack of net mean change obscures a significant amount of gross individual attitude change. About

Table 8: Multinomial logistic regression predicting the change in out-group feelings (T2 – T1) with clustered robust standard errors

	More negative			More positive		
	B	S.E.	Sign.	B	S.E.	Sign.
Intercept	–1.157	1.108	0.264	0.054	1.050	0.951
PDQI 1	–0.159	0.151	0.273	–0.209	0.109	0.051
PDQI 2	0.126	0.117	0.247	0.046	0.076	0.537
Group composition	–0.613	0.291	0.029	–0.027	0.230	0.950
Homogeneous						
Divided (ref.)						
Gender	0.112	0.197	0.291	0.100	0.237	0.648
Men						
Women (ref.)						
Age	0.002	0.009	0.760	0.011	0.009	0.189
Education	–0.151	0.362	0.673	–0.589	0.264	0.022
Lower or secondary						
Higher (ref.)						
Language	0.455	0.284	0.112	0.832	0.245	0.001
Dutch						
French (ref.)						

Notes: Ref. = 'No change'; Nagelkerke R^2 = 8.2%; n = 478.

37.7 per cent of participants did not report any change, whereas 31.5 per cent of them became more positive towards the out-group, and 30.6 per cent held more negative feelings after the deliberation. We thus see that the proportion of participants becoming more negative is larger than in the first study. This reflects the fact that the setting was hotter, and that there may have been a stronger ideological amplification on the negative side as a consequence (Sunstein, 2007).

To determine whether a high (perceived) deliberative quality and a divided group composition led to a more positive attitude towards the out-group, we ran the same analysis as in the first study. However, the normality assumption of the dependent variable was violated in this experiment. We therefore trichotomised the dependent variable into those who became more negative, those who did not change and those who became more positive. In Table 8 we ran a multinomial logistic regression predicting whether the participants became more positive or negative, with those who did not shift their opinion as the reference category. Because multinomial regression analyses are demanding in terms of cases, multi-level analyses would produce highly unstable results. We therefore once again clustered the standard errors to account for the fact that individuals are nested in groups.

First of all, we should look at the overall explained variance of the model. The R squared is relatively low, or at least much lower than in the previous study. In part, this can be explained by the use of a multinomial logistic regression, which inevitably leads to a loss of richness in

the data and reduces the overall variation in the dependent variable. But for the most part, we should keep in mind that the results of the first study were generated in a controlled laboratory setting, in which all potential confounders were identified and isolated. For the second study, the scale of the experiment was much larger, and the group dynamics were much harder to control. It was higher on mundane realism but lower on experimental realism. It is therefore more difficult to single out the crucial variables explaining the preference change, which leads to a lower explained variance.

The results in Table 8 hint at the same trends as in the first study, but the results are more nuanced. First, the group composition is significantly related to attitude transformations, but only for becoming more negative. Being confronted with the out-group, according to the results, can be considered a good buffer against becoming more negative. People seated at divided tables thus have lower odds of becoming negative than those seated at the unilingual tables. This finding suggests, in line with the first study, that direct confrontation and interaction with the out-group effectively reduces the chances of attitudes becoming more extreme towards the negative.

Second, the group composition is not significantly related to becoming more positive, but the perception of deliberation (PDQI 1) is. However, the direction of the effect is the inverse of what we would expect. Participants who have a high appreciation of their own role in deliberation are more likely to stand by their initial attitude. In an additional analysis (not shown here) we also added the interaction of both PDQI scales with the group composition to the multinomial logistic regression, but the interaction effects were not significant.

5 Discussion

In this article, we set out to offer an empirical perspective on the question of whether inter-group deliberation leads to out-group appreciation. More specifically, our aim was to find out whether a high quality of deliberation could foster changes in out-group attitudes, and whether deliberation in mixed company would improve out-group feelings. Our results are interesting, yet nuanced. First of all, both studies show that opinions do tend to shift, and they can even shift dramatically. However, contrary to expectations, the (actual or perceived) quality of the deliberation had only a limited impact upon attitude changes. The first study does not lend support to the first hypothesis that a high quality of deliberation leads to attitude changes, but we should take into account that the sample size was too low actually to disconfirm the hypothesis. And the second study finds only a limited effect of the perception of deliberation. This nuances the assumption often held by deliberative theorists that high-quality deliberation would be conducive to more open and magnanimous attitudes towards others. Moreover, the effects of deliberation did not interact with the group composition.

These findings mean that the transformative effects that deliberative scholars take largely for granted are not self-evident. The link between the process and outcomes of deliberation are not straightforward. In the wake of Maoz and Ellis's (2008) research on the mediating role of out-group trust, further research should try to unravel this intricate relationship, and – even though they are relatively modest – our findings do suggest that the way forward would be to focus on the perceived, rather than the actual, quality of deliberation.

The most important finding is, however, that most of the attitude change seems to be caused by the group composition. After all, the first study showed that confrontation with the

out-group is the best predictor of the shift towards a more positive out-group attitude, and in the second study, we found the participants at the homogeneous tables had a higher chance of becoming more negative towards the out-group. These findings offer further evidence that inter-group contact fosters inter-group appreciation, or at least that deliberation in an inter-group setting can operate as a buffer against more negative feelings. They even suggest that long and highly structured contact is not necessary for improving inter-group relations as the inter-goup contact hypothesis suggests (Pettigrew and Tropp, 2006). Rather, short communicative interactions might already exhibit some beneficial effects.

6 Conclusion

At first sight, deliberative democracy in divided societies seems doomed to fail. When deliberation brings together groups with opposing views entrenched in long-standing divisions, participants' willingness to yield to the force of the better argument is likely to be very low. But this article, in the wake of previous work on inter-group communication, shows the opposite dynamic: inter-group deliberation can actually lead to inter-group appreciation. It contributes to wider debates in political science and psychology by demonstrating that group composition matters more than deliberation quality. This is an important finding both for deliberative democracy and for the politics of accommodation in divided polities.

These findings call for future research in at least two directions. On the one hand, in order to go beyond the mere contact hypothesis, the content of the deliberation should be qualitatively explored to determine whether the effects are due to what was said or how it was said. Besides critical discourse analysis, the study of conceptual metaphors in citizen discourse constitutes a promising path to understanding how language can be used strategically (i.e. to sell one's argument) and pedagogically (i.e. to explain one's environment). Alternatively, experimental methods relying on different tasks between the participants of an inter-group deliberation could offer insight about the mechanisms behind the contact hypothesis.

On the other hand, the long-term impact of inter-group deliberation has still to be assessed. At the level of individuals, qualitative research based on interviews with the participants confirms that the participation in inter-group deliberation has a strong short-term impact; how it evolves over time is unknown. At the group level and at the level of society, the links between mini-publics and the maxi-public needs to be explored. While participants in mini-publics end up with a more positive view towards the out-group, a transfer of these new feelings to the rest of the population is unlikely to be automatic. This is an insightful and important area of investigation because it relates directly to the question of how to scale up deliberation.

Finally, even though we should be careful not to stretch our generalisations beyond the specific setting of these two Belgian experiments, our findings do suggest that the strategy of societal segmentation that is often applied in divided societies might do more harm than good. As such, our study offers hopeful perspectives for fighting the perpetuation of inter-group conflict by fostering inter-group deliberation.

About the Authors

Didier Caluwaerts is a Democracy Fellow at the Ash Center for Democratic Governance and Innovation of Harvard University, and a postdoctoral researcher of the Research Foundation-Flanders (FWO) at the Vrije Universiteit Brussel, Belgium. His research interests include deliberative democracy, political psychology and federalism. Didier Caluwaerts, Department of Political Science, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. E-mail: *dcaluwae@vub.ac.be*

Min Reuchamps is Professor of Political Science at the Université Catholique de Louvain, Belgium. His research interests are federalism and multi-level governance, political sociology and democratic innovations. Min Reuchamps, Institut de sciences politiques Louvain-Europe (ISPOLE), Centre de science politique et de politique comparée (CESPOL), Université Catholique de Louvain, Place Montesquieu 1, bte L02.08.07, 1348 Louvain-La-Neuve, Belgium. E-mail: *min.reuchamps@uclouvain.be*

Acknowledgements

The authors would like to thank André Bächtiger, Jürg Steiner, Juan Ugarriza and the three anonymous reviewers of this journal for their valuable comments on earlier versions of this article. The authors would also like to express their gratitude to Audrey Bisalu, Michiel Ceulemans, Belinda Drèze, Vittoria Fortuna, Benoît Lecloux, Elodie Martinet, Florence Mierzwa, Julie Reginster and Emilie Wattiau for their research assistance. This research was made possible with the generous help of the Research Foundation – Flanders (FWO), the PartiRep project, and the *Fonds spéciaux pour la recherche* of the Université Catholique de Louvain, Belgium.

Notes

1 In the divided groups simultaneous translation was provided to avoid any language bias.

2 Belgium consists of approximately 60 per cent Dutch speakers and 40 per cent French speakers. A disproportionately random sample (50–50 per cent) was used to ensure a balance between the linguistic groups in the experiments.

3 This panel is the largest in Belgium and was recently created on the occasion of a nationwide voting advice application organised by a consortium of universities.

4 Simultaneous translation was also undertaken at the divided tables.

5 Ten per cent of the places were nonetheless reserved for participants not recruited randomly. Despite the careful process of random selection, there was a stronger dropout among the groups that are traditionally less interested in the subject. This is why 10 per cent of them were recruited through intermediary social organisations which have bonds of trust with these underprivileged groups (Ryfe, 2005).

6 For the operationalisation of these two additional variables, see Caluwaerts, 2012.

7 We reran the analyses with the thricotomisation of the education variable, and the results proved robust.

8 Because of the small number of cases at the group level, MLA was not possible.

References

Allport, G. (1954) *The Nature of Prejudice*, Reading, MA: Addison-Wesley.

Black, L., Burkhalter, S., Gastil, J. and Stromer-Galley, J. (2011) 'Methods for Analyzing and Measuring Group Deliberation' in E. Bucy and R.L. Holbert (eds.), *The Sourcebook of Political Communication Research*, New York: Routledge, pp. 323–346.

Bogaards, M. (1998) 'The Favourable Factors for Consociational Democracy: A Review', *European Journal of Political Research* 33(4), pp. 475–496.

Caluwaerts, D. (2012) *Confrontation and Communication: Deliberative Democracy in Divided Belgium*, Brussels: Peter Lang.

Caluwaerts, D. and Reuchamps, M. (2012) *The G1000: Facts, Figures and Some Lessons from an Experience of Deliberative Democracy in Belgium*, Brussels: Re-Bel Initiative. Available from: http://www.rethinkingbelgium.eu/rebel-initiative-files/events/seventh-public-event-g1000-european-citizens-initiative-malaise-democracy/G1000-Background-Paper.pdf [Accessed 17 December 2013].

Caluwaerts, D. and Ugarriza, J.E. (2012) 'Favorable Conditions to Epistemic Validity in Deliberative Experiments: A Methodological Assessment', *Journal of Public Deliberation* 8(1), pp. 1–20.

Cappella, J.N., Price, V. and Nir, L. (2002) 'Argument Repertoire as a Reliable and Valid Measure of Opinion Quality: Electronic Dialogue during Campaign 2000', *Political Communication* 19(1), pp. 73–93.

Chambers, S. (2003) 'Deliberative Democratic Theory', *Annual Review of Political Science* 6, pp. 307–326.

Cohen, J. (1960) 'A Coefficient of Agreement for Nominal Scales', *Educational and Psychological Measurement* 20(1), pp. 37–46.

Delli Carpini, M.X. and Keeter, S. (1996) *What Americans Know about Politics and Why It Matters*, New Haven, CT: Yale University Press.

Deschouwer, K. and Reuchamps, M. (2013) 'The Belgian Federation at a Crossroad', *Regional and Federal Studies* 23(3), pp. 261–270.

Dovidio, J.F., Gaertner, S.L. and Kawakami, K. (2003) 'Inter-group Contact: The Past, Present, and the Future', *Group Processes and Inter-group Relations* 6(1), pp. 5–21.

Dryzek, J.S. (2000) *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*, Oxford: Oxford University Press.

Ellis, D.G. and Maoz, I. (2002) 'Cross-Cultural Argument Interactions between Israeli-Jews and Palestinians', *Journal of Applied Communication Research* 30(3), pp. 181–194.

Ellis, D.G. and Maoz, I. (2007) 'Online Argument between Israeli Jews and Palestinians', *Human Communication Research* 33(3), pp. 291–309.

Fiket, I., Olsen, E.D.H. and Trenz, H.-J. (2011) *Deliberation under Conditions of Language Pluralism: Insight from the Europolis Deliberative Polling Experiment*, Oslo: Arena.

Fishkin, J.S. (1995) *The Voice of the People: Public Opinion and Democracy*, New Haven, CT: Yale University Press.

Fung, A. (2003) 'Recipes for Public Spheres: Eight Institutional Design Choices and Their Consequences', *Journal of Political Philosophy* 11(3), pp. 338–367.

G1000 (2012) *Final Report: Democratic Innovation in Practice*. Available from: http://www.g1000.org [Accessed 25 October 2013].

Gastil, J., Bacci, C. and Dollinger, M. (2010) 'Is Deliberation Neutral? Patterns of Attitude Change during "The Deliberative Polls" ', *Journal of Public Deliberation* 6(1), pp. 1–33.

Hewstone, M., Rubin, M. and Willis, H. (2002) 'Inter-group Bias', *Annual Review of Psychology* 53, pp. 575–604.

Hickerson, A. and Gastil, J. (2008) 'Assessing the Difference Critique of Deliberation: Gender, Emotion, and the Jury Experience', *Communication Theory* 18(2), pp. 281–303.

Holsti, O.R. (1969) *Content Analysis for the Social Sciences and Humanities*, Reading, MA: Addison-Wesley.

Hooghe, M. (1999) 'The Rebuke of Thersites: Deliberative Democracy under Conditions of Inequality', *Acta Politica* 34(4), pp. 287–301.

Karpowitz, C.F., Mendelberg, T. and Shaker, L. (2012) 'Gender Inequality in Deliberative Participation', *American Political Science Review* 106(3), pp. 533–547.

Luskin, R.C., O'Flynn, I., Fishkin, J.S. and Russell, D. (2012) 'Deliberating across Deep Divides', *Political Studies*, Early View. DOI: 10.1111/j.1467-9248.2012.01005.x

Maoz, I. and Ellis, D.G. (2008) 'Inter-group Communication as a Predictor of Jewish-Israeli Agreement with Integrative Solutions to the Israeli–Palestinian Conflict: The Mediating Effects of Out-Group Trust and Guilt', *Journal of Communication* 58(3), pp. 490–507.

Mendelberg, T. (2002) 'The Deliberative Citizen: Theory and Evidence' in M.D.D. Carpini, L. Huddy and R.Y. Shapiro (eds.), *Research in Micropolitics: Political Decisionmaking, Deliberation and Participation*, Greenwich, CT: JAI Press, pp. 151–193.

Meyers, R.A., Brashers, D.E. and Hanner, J. (2000) 'Majority–Minority Influence: Identifying Argumentative Patterns and Predicting Argument–Outcome Links', *Journal of Communication* 50(4), pp. 3–30.

Morrell, M. (2010) *Empathy and Democracy: Feeling, Thinking and Deliberation*, University Park, PA: Penn State University Press.

Mutz, D. (2006) *Hearing the Other Side: Deliberative vs. Participatory Democracy*, Cambridge: Cambridge University Press.

O'Flynn, I. (2006) *Deliberative Democracy and Divided Societies*, Edinburgh: Edinburgh University Press.

Pettigrew, T.F. and Tropp, L.R. (2006) 'A Meta-analytic Test of Inter-group Contact Theory', *Journal of Personality and Social Psychology* 90(5), pp. 751–783.

Reuchamps, M. (2013) 'The Future of Belgian Federalism through the Eyes of the Citizens', *Regional and Federal Studies* 23(3), pp. 353–368.

Reuchamps, M., Kavadias, D. and Deschouwer, K. (2013) 'Drawing Belgium: Using Mental Maps to Measure Territorial Conflict', *Territory, Politics, Governance* [online]. DOI: 10.1080/21622671.2013.814549

Ryfe, D.M. (2005) 'Does Deliberative Democracy Work?', *Annual Review of Political Science* 8(1), pp. 49–71.

Steenbergen, M.R., Bächtiger, A., Spörndli, M. and Steiner, J. (2003) 'Measuring Political Deliberation: A Discourse Quality Index', *Comparative European Politics* 1(1), pp. 21–48.

Steiner, J. (2012) *The Foundations of Deliberative Democracy*, Cambridge: Cambridge University Press.

Sunstein, C.R. (2007) 'Ideological Amplification', *Constellations* 14, pp. 273–279.

Wojcieszak, M. (2012) 'On Strong Attitudes and Group Deliberation: Relationships, Structure, Changes, and Effects', *Political Psychology* 33(2), pp. 225–242.