Examining the 'Gold Standard':
Deliberative Quality in the Transnational Deliberative Poll Europolis

Inauguraldissertation zur Erlangung der Würde eines Doctor rerum socialium der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Bern

vorgelegt von
Marlène Gerber
von Langnau i.E., Bern

2013

Die Fakultät hat diese Arbeit am 23. Mai 2013 auf Antrag der beiden Gutachter Prof. Dr. Marco Steenbergen und Prof. Dr. Michael A. Neblo als Dissertation angenommen, ohne damit zu den darin ausgesprochenen Auffassungen Stellung nehmen zu wollen.

Contents

List of Tables

List of Figures

Acknowledgments

I would like to sincerely thank my supervisors Marco Steenbergen (University of Zurich) and Michael Neblo (Ohio State University). Their insightful comments and suggestions on previous versions of these papers have considerably influenced my analytic thinking and have helped to locate my research in the growing and highly diverse field of deliberative democracy. In addition, I owe my deepest gratitude to André Bächtiger (University of Lucerne). His guidance into the world of deliberation, and into the academic world in general, has been essential during the three years of my PhD and has strongly shaped my understanding of what it means (and takes) to become a researcher. I am also profoundly indebted to Simon Reber, Samuel Rohr and Mirjam Rütsch. They showed great motivation in their careful assistance of transcribing and coding the group discussions, as if the thesis to be written was their own. In addition, I would like to thank my colleague Seraina Pedrini. I have benefitted greatly from our discussions on deliberative democracy and highly valued the constant scientific exchange that we kept over the last three years. Special thanks go as well to Jürg Steiner (University of Chapel Hill), whose unwavering motivation for deliberation was very inspiring from the start. I would also like to express my gratitude to Marc Bühlmann (University of Bern). His scientific advice and constant reminder that the thesis comes first was highly appreciated, particularly given that he employed me for a project unrelated to my PhD. Furthermore, it is a pleasure to thank the researchers at the Center of Excellence on Democracy Research at Åbo Akademi, Finland, who kindly hosted me for two months at the final stage of my dissertation. The inspiring research atmosphere and the great hospitality rendered Åbo/Turku the ideal place to be at that time.

Various versions of my PhD papers have been presented to numerous people at conferences and seminars. I would like to thank all of them for their comments, suggestions and encouragements. In one way or another, they all contributed to shaping the ideas of the topic under investigation and a considerable number gave me pause to think, write and re-write many of the passages now included in this thesis. It goes without saying that the remaining mistakes are the author's.

I owe my thanks to the organizers of Europolis, notably James Fishkin (University of Stanford) and Pierangelo Isernia (University of Siena), as well as to all the motivated citizens, without whose participation the Europolis event could have never been realized. This thesis

would not have been possible without the financing of the Swiss National Science Foundation. I would like to express my sincere thanks for this three-year project funding. In addition, I am grateful to the Faculty of Economics and Social Sciences at the University of Bern and the Swiss Academy of Humanities and Social Sciences (SAGW), whose research travel grants made the participation at various international conferences possible.

In addition, I would like to thank my parents, Ruth and Bernhard, for their constant and indispensable moral support. I would also like to thank my friends for offering welcome distraction at those times when the PhD project risked turning into a 24/7 endeavor. And last but not least, I would like to offer my special thanks to my partner, Pascal, for his admirable patience and endless personal support along the way, particularly in its final stage when the project definitely became a 24/7 endeavor. By just being there, these people contributed substantially to the realization of my thesis.

Marlène Gerber,

Thun, May 2013

Chapter 1

INTRODUCTION

Two decades ago, deliberative democracy emerged as a normative theory from an ideal of democratic legitimacy (e.g., Bohman, 1998; Chambers, 2003; Habermas, 1996). At the core of deliberative democracy lies the premise of political justification via public reasoning among free and equal subjects (e.g., Cohen, 1996; Elster, 1998; Gutmann and Thompson, 2004; Mansbridge et al., 2012). In contrast to a mere aggregation of fixed preferences through voting, deliberation offers a talk-centric model of democracy in which decision-making is preceded by communicative interaction between subjects, who are required to adapt their preferences in the face of new evidence and superior reasoning (e.g., Chambers, 2003; Elster, 1997; Knight and Johnson, 1994). Building on an ideal notion of political discourse, deliberative democracy theory had to face considerable critique from opponents who questioned its applicability to real-world settings (e.g., Eliasoph, 1998; Hibbing and Theiss-Morse, 2002; Posner, 2004). For a long time, however, the discussion remained rather theoretical and culminated in a stalemate between "a priori skepticism and untested idealism" (Bohman, 1998: 422). Empirical research was subsequently proposed as a way to move the debate forward (Bohman, 1998). As a result, the past decade has experienced a strong surge in empiricism, moving deliberative democracy from a normative theory to a stage of "working theory" (Chambers, 2003; see Delli Carpini, et al. 2004; Mutz, 2008; Thompson, 2008 for reviews). While instances of deliberation have been detected in the national (e.g., Bara et al., 2007; Esterling, 2011; Mucciaroni and Quirk, 2006; Pedrini et al., 2013; Steiner et al., 2004; Weale et al., 2012) and supranational (e.g., Doerr, 2009; Holzinger, 2001; Naurin, 2010; Ulbert and Risse, 2005) political sphere, tremendous importance has been attributed to tracking deliberation in the public sphere (e.g., Smith, 2011). On the one hand, deliberation has been studied in real-world settings such as everyday talks (e.g., Conover et al., 2002; Conover and Searing, 2005; Price et al., 2002), town meetings (e.g., Bryan, 1999; Mansbridge, 1983; Mendelberg and Oleske, 2000; Rao and Sanyal, 2009), and online forums (e.g., Graham and Witschge, 2003; Janssen and Kies, 2005; Papacharissi, 2004). On the other hand, artificial settings of interaction have been created in order to monitor citizen deliberation, in what have been termed mini-publics, i.e. in "more perfect public spheres" (Fung, 2003: 338). By gathering a cross-section of the population and by incorporating certain procedural safeguards such as facilitators and rules for civil discussion, mini-publics attempt to provide a safe venue for deliberation that is free from "certain negative or distorting effects of the broader public

sphere" (Chambers, 2004: 400).[1] The idea that mini-publics may render individuals' opinions more informed and considered spread quickly and resulted in a global proliferation of various forms of mini-publics (e.g., Dryzek, 2009; Fishkin et al., 2010; Grönlund et al., 2010; Smith and Wales, 2000; Warren and Pearse, 2008). The exponential increase in democratic innovations resulted in a shift from *deliberative democracy* to *democratic deliberation*, translating the hopes of deliberative democrats from the broader public sphere to small-scaled, self-contained initiatives of artificial public deliberation (Chambers, 2009; Mansbridge, 2007).

Among these artificial settings of public deliberation, the deliberative poll (DP) developed by James Fishkin (1995) is probably the most widely known mini-public and has been applied to a variety of national and transnational settings (e.g., Andersen and Hansen, 2007; Fishkin, 2009; Fishkin et al., 2010; Luskin et al., 2002; Luskin et al., 2013).[2] Jane Mansbridge (2010) recently praised DP as the *gold standard* among contemporary mini-publics. DP possesses many of the favorable features for successful deliberation (e.g., Fishkin and Luskin, 2005; Mansbridge, 2010). Most importantly, these include the gathering of a statistically representative sample, the provision of balanced information material, the possibility to question experts and politicians on various policy alternatives, and the presence of trained facilitators who should ensure that group discussions are balanced and civil. Deliberative polls are also meant to "insulate people from social pressure" by depriving participants from reaching a decision, a common statement, or any other kind of mutual agreement that could distort the process of uncoerced argumentation (Ackerman and Fishkin, 2002: 134; see also Sunstein, 2002).

Positive evaluations have led to DP being viewed as the model for successful public deliberation. Compared to other mini-publics, DP usually scores highest on the input-oriented criteria of representation (e.g., Fung, 2003). When it comes to output-oriented criteria, evaluations show that participants often change their opinion, gain knowledge and report high levels of satisfaction with participating during the event (e.g., Andersen and Hansen, 2007; Fishkin et al., 2010; Luskin et al., 2002). Although insights on these input- and output-criteria have been rich, the *process* of deliberation in DP has largely remained unexamined (for an exception, see Siu, 2009). Despite the fact that the body of empirical research on deliberation is growing, attempts to systematically investigate what actually happens when people

[1] For a review on existing mini-publics, see Fung (2003), Goodin and Dryzek (2006), and Smith (2008, 2009). Fung's definition of mini-public is somewhat broader in that it also includes self-selected settings or settings where deliberators have been elected by a special body.

[2] See http://ccd.stanford.edu for more information on the conduct of various deliberative polls.

deliberate in mini-publics have, until recently, been scarce.[3] Too often, conclusions are based on the notion of idealized actors who successfully deliberate in the absence of any distorting contextual effects. Or, as Ryfe (2005: 54) puts it: "Researchers have been less interested in deliberation itself than in measuring its effects". Strikingly, just as deliberative theory initially cared little about empirical verification, empirical research in deliberation seemed to care little about theoretical falsification. This is a significant shortcoming, given that available insights from real-world deliberation and the psychological literature on group processes raise important doubts about the possibility of biased-free communication among ordinary citizens (e.g., Conover and Searing, 2005; Mansbridge, 1983; Mendelberg and Oleske, 2000; Sanders, 1997; see Delli Carpini, et al. 2004; Lupia, 2002; Mendelberg, 2002; Mutz, 2008 for reviews on the psychological literature). Of course, drawing far-reaching implications from settings that lack supportive institutional underpinnings for successful deliberation may not be appropriate. Still, we should be sure that what is seen as the role model for citizen deliberation actually delivers what it promises. In response, various reservations have been aired that all share the same underlying question: "How are we supposed to know, if we don't know?" (e.g., Azmanova, 2010; Lupia, 2002; Mutz, 2006; Price and Neijens, 1998; Sanders, 2010). How can we know that the favorable institutional setting was sufficient to establish a balanced, civil and considerate atmosphere for discussion, if we actually don't know? And how can we be sure that post-deliberative opinions are the result of a careful reasoning process instead of the mere provision of additional information or, what would be worse, the consequence of manipulation?

The aim of this thesis is to expose a recent deliberative poll to this critique. In doing so, the thesis attempts to provide a systematic and in-depth analysis of the deliberative process at hand. It seeks to advance a conception of deliberation that is strongly rooted in normative theory but at the same time accounts for practical limitations of real-world deliberation. Concretely, the thesis addresses the question of whether and to what extent the deliberative poll under examination lives up to the promises of deliberative theory, within these practical limitations. For an empirical example, the thesis examines a subsample of small group discussions in Europolis, a pan-European deliberative poll on third country migration and climate change that was held in May 2009 in Brussels (Fishkin, 2009).[4] By including participants from all 27 EU-member states, Europolis provided an extensive attempt for cross-

[3] Exceptions include e.g., Stromer-Galley (2007); Siu (2009); Himmelroos (2011); Steiner (2012); Pedrini et al. (2013).

[4] See Chapter 2 for further information on Europolis and the selected groups.

cultural deliberation in a multilingual setting. Organized in chapters, the thesis consists of the three central pieces of work of my cumulative dissertation. These address the following issues of deliberative quality in Europolis:

Chapter 2, a collaboration with André Bächtiger, Susumu Shikano, Simon Reber, and Samuel Rohr, investigates the deliberative abilities of ordinary citizens. Drawing from a philosophically grounded measure, an updated version of the Discourse Quality Index (DQI) (Bächtiger et al., 2010b; see Steenbergen et al., 2003), it explores to what extent ordinary European citizens are capable of meeting deliberative ideals and whether socio-economic, cultural, and psychological biases show up in the ability to deliberate. By incorporating a measure that tabs whether participants use personal narratives or experiences to back their claims, the updated DQI accounts for recent developments in deliberative theory that call for expanding the admissible forms of deliberative communication (Bächtiger et al., 2010a). The chapter also pursues a strongly methodological perspective and applies a novel conceptualization of deliberative quality. Applying a Bayesian item response analysis allows us to explore how well citizens are able to achieve the various standards of deliberative quality (justification rationality, common good orientation, empathy, respect, story-telling and questioning). Moreover, it allows for an assessment of whether the various standards actually map an underlying construct of deliberative quality by simultaneously holding the realistic assumption that some standards of deliberative quality may be more difficult to achieve than others. The major finding is that while true deliberators scoring high on all deliberative components exist in the realm of citizen deliberation, deliberative skills are unequally distributed. Working-class participants from Central Eastern and, to a lesser degree, from Southern Europe displayed a lower deliberative competence than other participants. The chapter also makes a first attempt to explore whether deliberative abilities are linked to the propensity to change opinions. Findings reveal that both the most and the least skilled deliberators change their opinions to a similar degree.

With a focus on equality, Chapter 3 is dedicated to another crucial standard of deliberation. The chapter takes up the claim of difference democrats who argue that informal hurdles to participation can persist even after individuals gain formal access to the political arena (e.g., Fraser, 1992; Sanders, 1997; Young, 2002). I argue for an extended approach to deliberative inclusiveness that goes beyond evaluating the representativeness of the participant sample. Specifically, I assess whether participants contributed in roughly equal measures to the discussion and whether their discussion partners considered their contributions on equal merits. In doing so, the chapter adds to the small but growing literature on deliberation that

expresses reservations about taking the willingness to engage with others claims for granted (e.g., Burkhalter et al., 2002; Dobson, 2010, 2012; Pedrini et al., 2013; Stromer-Galley, 2007). In order to account for the intrinsically relational aspect of interpersonal communication, measures of social network analysis are introduced as possible tools to evaluate the inclusiveness of deliberative encounters (see e.g., Cinalli and O'Flynn, 2013; Graham and Witschge, 2003). Findings indicate that, despite the presence of trained facilitators, some traditionally disadvantaged groups contribute significantly less to the discussion in terms of speaking time. However, concerns are somewhat reduced in relation to consideration. In addition, equality seems to be particularly at risk in groups that display a considerable degree of national heterogeneity. The latter finding strongly encourages further investigation of contextual determinants for successful deliberation.

At the core of deliberative polling is the question of whether participants change their minds after deliberation. From a normative point of view, post-deliberative opinions should be linked to the quality of arguments presented during discussion. Yet, there is a dearth of research testing this claim. The 4th and last chapter of my thesis makes a first attempt to overcome this deficiency (co-authored with André Bächtiger, Irena Fiket, Marco Steenbergen, and Jürg Steiner). Specifically, we explore whether statements backed by reasons affect opinions, which we term deliberative persuasion. We contrast deliberative persuasion to non-deliberative persuasion, whereby we explore whether the most frequently repeated rather than the most justified position influences opinions. We focus on two crucial questions that were discussed in Europolis small group discussions on immigration: the regularization of irregular immigrants and the European involvement in immigration affairs. We find some evidence that, in the context of regularization of irregular immigrants, deliberative persuasion took place. Yet this result does not reflect a general trend: in the context of European involvement in immigration affairs, post-deliberative opinions are driven by the most frequently repeated position rather than by argumentative rationality.

Too often, outcomes of mini-publics have not been sufficiently embedded in the larger deliberative system (e.g., Chambers, 2009; Goodin and Dryzek, 2006; Olsen and Trenz, 2013; Parkinson, 2006b, 2006c; but see Fishkin and Luskin, 2006; Fishkin et al., 2010). Neither does the thesis aspire to do so. However, I believe that one potential pathway to overcome this 'scaling up' problem is an approach that relies on scientific discoveries. A systematic evaluation of the deliberative process in various mini-public designs may allow for the determination of the most favorable conditions for successful deliberation. In the long run, such an approach may increase credibility and trust in existing (and improved) mini-public

designs and their consequences. In the end, this may allow mini-publics to grow out of their experimental state and become a viable complement to existing participatory processes in modern democracies.

Chapter 2

DELIBERATIVE ABILITIES OF ORDINARY CITIZENS*

Marlène Gerber, André Bächtiger, Susumu Shikano, Simon Reber, Samuel Rohr

* We would like to thank John Dryzek for his encouragement and support. We are also grateful to the participants of the workshop "Frontiers of Deliberation" at the ECPR Joint Sessions of Workshops in St. Gallen 2011. Finally, we would like to thank Karolina Kojder for her careful work in transcribing and translating the Polish group discussions.

2.1 Introduction

Can ordinary citizens deliberate together at high quality levels? By high quality deliberation, we mean that participants in deliberation present reasons for their positions, make reference to the common good, are respectful, empathetic, and inquisitive (see Mansbridge et al., 2010). For James Fishkin, one of the deliberative pioneers, the lessons from the manifold experiences with citizen deliberation worldwide are clear: everybody can deliberate (e.g., Ackermann and Fishkin, 2002; Fishkin and Luskin, 2005). Fishkin is not the only one claiming that ordinary citizens can be turned into good deliberators. The thrust behind the deliberative movement is that Schumpeterian conceptions of the minimalist democratic citizen are woefully wrong and that with a little helping hand, ordinary citizens can approach philosophical ideals of deliberation, even in a transnational context (see, e.g., Dryzek, 2010). Not surprisingly, several proposals for a deliberative involvement of citizens in global affairs have been forwarded in recent years, claiming that deliberating citizens may do a better job than elected politicians in finding solutions and common ground in frequently deadlocked global politics (Dryzek et al., 2011; Goodin and Ratner, 2011).

But the deliberative gospel has not convinced everyone. Since the advent of deliberative theories in the 1990s, deliberation has met with sustained criticism. Critiques have revolved around deliberation's inconsistency with psychological theories and experiences of human action (e.g., Mutz, 2008; Posner, 2003). Indeed, when deliberative advocates refer to deliberation's success in practice, the reaction of many empirical political scientists is negative. As Mutz (2008: 533) writes: "As an empirical theory, deliberative theory has been widely criticized for making assumptions that seem to fly in the face of what scholars already know about human behavior". Critics of deliberation question the idealized conjectures of the "deliberative citizen" possessing sophisticated reasoning skills while being simultaneously respectful, reflective, inquisitive and open-minded. Drawing from psychological experiments and jury deliberations critics also claim that deliberative abilities are strongly biased along socio-economic, cultural, and psychological factors (see Devine et al., 2001; Krupnikov et al., 2007; Mendelberg, 2002). If such biases in deliberative ability exist - and if they also transfer to the outcomes of deliberation - then the deliberative ideal of an egalitarian and unconstrained exchange of arguments is violated.

So far, however, a great deal of this controversy is surprisingly theoretical rather than based on systematic empirical facts. On the one hand, many critics tend to overlook that modern-day

citizen deliberation is a highly structured affair. Most citizen deliberations are conducted under supportive institutional conditions, i.e., citizens get balanced information material, experts answer citizens' questions, and facilitators ensure that small group discussions are on-topic and focused on all arguments. Thus, drawing far-reaching implications from psychological experiments and jury deliberations that lack such supportive institutional underpinnings may not be appropriate. On the other hand, deliberationists tend to assume that reasoned discourse will quasi-automatically follow when conditions for deliberation are good. But they have downplayed the possibility that even under optimal conditions, deliberation may not occur as expected by theory. Not only may some citizens be overtaxed in the deliberative process, citizen "deliberation" may also be just talk, absent any philosophical underpinnings. In the past decade, several scholars have tried to address the objections against deliberation's viability via empirical research. This empirical defense, however, is mostly concerned with the input dimension, i.e., the inclusiveness of the deliberative process, and the output dimension, i.e., the question whether deliberation leads to biased opinion and knowledge changes (e.g., Barabas, 2004; Conover et al., 2002; Cook et al., 2007; Luskin et al., 2002; Sturgis, 2003). The actual process of deliberation, however, remains understudied (see Ryfe, 2005). Yet, much punch behind deliberative theory is that preferences are constructed during the deliberative process (Parkinson, 2012), requiring that we analyze deliberative abilities and potential biases in that process as well. While recent studies have included measures of process quality, this is mostly done via survey-based self-reports of participants (e.g., Andersen and Hansen, 2007). But this approach is deficient as well: not only may self-reports contain elements of social desirability; they may also insufficiently capture the philosophical ramifications of the deliberative model. In sum, both critics and advocates of deliberation have barely made the effort of analyzing what citizens actually do when they are supposed to deliberate (among the few exceptions are Dutwin, 2003; Himmelroos, 2011; Pedrini, 2013; Schneiderhan and Kahn, 2008; Siu, 2009).

Our article sets out to change this unhappy state of affairs. We focus on actual deliberative behavior of citizens rather than on potentially biased survey-based evaluation. Moreover, we study deliberative behavior in the context of supportive institutional conditions, namely deliberative polling (Fishkin and Luskin, 2005). Concretely, we focus on *Europolis*, a pan-European deliberative poll on third country migration and climate change, which was carried out in Brussels in late May 2009. We explore to what extent ordinary European citizens were able to meet deliberative ideals and whether socio-economic, cultural, and psychological biases showed up in the ability to deliberate. To measure deliberative quality, we employ an

updated version of the *Discourse Quality Index* (Steenbergen et al., 2003) which employs a broad understanding of deliberative quality and allows for a quantitative content analysis of recorded discussions. In order to determine what counts as "high" and "low" deliberative quality, we draw from a classic and Habermasian-inspired understanding of deliberative quality (which we adapt to the constraints of citizen deliberation, however). The aggregation of the different DQI components is accomplished via Bayesian *Item Response Analysis*. In doing so, we explore how well citizens are able to achieve the various standards of deliberative quality (justification rationality, common good orientation, or respect), while simultaneously making the realistic assumption that some components of deliberative quality may be more difficult to achieve than others. By mapping all participants as ideal points on latent dimensions of deliberative quality, item response analysis enables us to check whether deliberative ability is biased along socio-economic, cultural, and psychological factors. Finally, we put the findings of our bias analysis in perspective by relating our measure of deliberative ability to another crucial measure of deliberative quality (and bias), namely participation equality, and by considering some consequences of deliberative ability, namely whether biases in deliberative ability also translate into differential opinion shifts.

By examining actual deliberative behavior of citizens in a supportive environment for fruitful deliberation, and by combining this with advances in measuring deliberation, we hope to move the controversy between advocates and critics of deliberation to a new level of sophistication. Notice that when we speak of deliberative ability we do not see this as a fixed and unchangeable trait, but as a *communicative repertoire* that some people may be better at than others but that every individual can eventually cultivate and develop.

The remainder of the article is organized as follows. Section 2.2 gives more background on the controversy between critics and advocates of deliberation and lays some theoretical groundwork. Section 2.3 is concerned with the measurement of deliberation: it presents an updated version of the Discourse Quality Index, defines cut levels for high quality deliberation, and introduces the aggregation method, Bayesian Item Response Analysis. Section 2.4 gives some background of the object of our study, the pan-European deliberative poll Europolis, and describes the data. Section 2.5 presents the empirical findings. Section 2.6 puts our findings in perspective and discusses the implications of our findings for a deliberative involvement of citizens in transnational and global affairs and the design of citizen deliberations. Section 2.7 concludes.

2.2 Deliberative Abilities: A Theoretical Overview

For quite some time, many deliberative theorists have presented a view of the deliberative citizen that is in line with the view of the democratic citizen offered by John Rawls (1993). As Rosenberg (2005: 213) holds: "rather than offering an empirical psychology of individual citizenship, Rawls stipulated what democratic theory requires". In a Rawlsian view, citizens must have the basic cognitive capacity to argue with reasons, be logical, and reflect on their own presuppositions (Rosenberg, 2005: 213–214). By contrast, many psychologists and skeptics of deliberation have argued that only a small minority of individuals possesses a level of deliberative ability as required by deliberative theory and that these abilities are also biased along socio-economic and other factors (e.g., Mendelberg, 2002; Rosenberg, 2005). As Sanders (1997: 348) has succinctly put it: "some citizens are better than others at articulating their arguments in rational, reasonable terms". Since a key goal of deliberation is to include 'all affected' and empower the disenfranchised, a tension between inclusion and deliberative ability – or what Young (2002: 38) aptly calls "articulateness" - may arise, turning deliberation into a potentially harmful intervention that further marginalizes already disenfranchised groups.

Even the most ardent deliberative theorists now acknowledge that citizens' deliberative abilities may vary considerably. Nonetheless, deliberationists still think that under favorable institutional conditions, a considerable number of citizens can approach deliberative ideals. In the following, we discuss the psychological foundations of a genuine deliberator and then turn to socio-economic, cultural, and psychological biases in deliberation.

2.2.1 Foundations of the Genuine Deliberator

From a philosophical perspective, an ideal deliberator must fulfill demanding behavioral obligations: she must be reasoned, common-good oriented, inquisitive, reflective, respectful, empathetic, and open to the better argument (see, e.g., Mansbridge et al., 2010). As mentioned before, a vast number of cognitive research in psychology suggests that individuals are not fully logical and therefore not likely to be fully rational in the manner specified by deliberative theory (see, e.g., Krupnikow et al., 2007; Reykowski, 2006; Rosenberg, 2005). Moreover, drawing from personality psychology (see, e.g., McCrae and Costa, 1987) it is not clear whether a genuine deliberator as envisioned by philosophers can exist in reality (Jennstal, 2012: 17). According to Jennstal (2012), an ideal deliberator must score high on at least three personality traits, namely extraversion, agreeableness, and openness. Hereby, reason-giving is a skill that primarily falls under the rubric of extraversion. Reflexivity,

respect, empathy, and inquisitiveness, in turn, require openness and/or agreeableness. From a personality trait perspective, a genuine deliberator represents a rare constellation of personality traits (Jennstal, 2012: 18). For instance, being extraverted does not necessarily mean that one also scores high on agreeableness and open-mindedness. Quite to the contrary, people who can reason well may not be very respectful and empathetic, even though complex thinking styles may coincide with inquisitiveness and openness (see Zhang and Huang, 2001). In other words, personality psychology questions the possibility of a "unitary" deliberator scoring high on all quality standards of deliberation. Unfortunately, the Europolis questionnaire does not include personality questions, thus we cannot directly link personality traits to deliberative behavior. But we can indirectly explore whether genuine deliberators exist, by checking whether the different components of deliberative quality – justification rationality, respect, empathy, and inquisitiveness – form a compound and coherent phenomenon at the level of deliberating citizens. Assume that a genuine deliberator scoring high on all deliberative standards indeed exists. Yet, we may still be confronted with the fact that this only represents a tiny minority of individuals who are also biased in socio-economic and other regards (see Rosenberg, 2005: 221). This is the topic of the next subsection.

2.2.2 Biases in Deliberation

There are myriads of potential biases in the ability to deliberate. In this article, we concentrate on a number of socio-economic, cultural, and psychological biases that we can capture on the basis of the Europolis questionnaire.

Socio-economic biases in deliberation

Difference democrats and feminists have claimed that certain social groups are disadvantaged in deliberation, namely women, blacks, poor, and less-educated people (Sanders, 1997; Young, 2002). The argument is that rational argumentation including logical deduction and general principles is frequently associated with men and socially privileged groups, while more tentative, figurative, and emotional forms of expression are often associated with women, socially less privileged groups, and cultural minorities (Young, 2002: 38–40). Regarding gender, some scholars have argued that women have higher capacities for respect and empathy and thus may even be the better deliberators than men (Mansbridge, 1996a: 123; Norris, 1996: 91). Regarding class, sociological research has argued that there may be class-specific ways of speaking and arguing. According to Bernstein (1971–1990), working-class people tend to adopt a restricted code of speech by using rather simple, repetitive and limited vocabulary that stands in contrast to the more accurate and elaborate code of speech

commonly employed by the middle-class. Class differences in the ability of speaking and arguing are closely linked to differences in education. Working-class people are embedded in an oral culture and thus lack 'literacy' writ large. By contrast, higher social classes and well-educated people are more likely to have deliberative skills. Education increases verbal cognitive proficiency, leads to improved comprehension of politics and governance and enhances tolerance towards opposing viewpoints (Nie et al., 1996). What is more, well-educated individuals have access to occupations where they can further develop their reasoning and public speaking skills and this situation plays its part in widening the gap in deliberative skills over time (see Mendelberg, 2002). Finally, age can be seen as a proxy for experience with political affairs (see Spörndli, 2004). Experience may increase a person's ability for self-reflection and responsiveness to others (Rosenberg, 2007).

Cultural biases in deliberation

Theorists of multiculturalism worry that classic deliberative ideals – such a rational argumentation - represent a culturally and historically specific format of communication (see, e.g., Young, 2002). In our sample of European citizens, this mainly concerns cultural dividing lines between Southern, Central Eastern, and Western Europeans. With regard to Southern Europeans, Gambetta (1998) has made a highly controversial argument, namely that "indexical" or "Claro!" cultures – an alleged feature of Southern European culture - are "antithetic to turn taking, mutual respect, sincerity, truthfulness and everything that contributes to successful deliberation" (Sass, 2006: 2). With regard to Central Eastern Europeans (CEE), several studies indicate that citizens from CEE countries have different democratic values than citizens from Western European countries: they are less engaged in civic affairs and politics and also have less trust in others (Fuchs and Klingemann, 2002; Karp and Banducci, 2007). This cultural dividing line may also transfer to different deliberative behavior, with the expectation that Central Eastern Europeans might have lower deliberative performance than Western Europeans. From an empirical vantage, several researchers have identified traces of cultural differences in deliberative behavior, finding that some societal cultures are less compatible with deliberative ideals (e.g. Duchesne and Haegel, 2007; Min, 2009). Yet alleged cultural differences may sometimes merely reflect different experiences with deliberative practices. Focusing on elite deliberation in working groups of the European Council of Ministers, Naurin (2010) finds that new member states from CEE displayed lower levels of deliberative quality than old member states. According to Naurin (2010: 47), this has to do with experience rather than culture, since old member states are "more accustomed at how to 'play the Brussels game'".

Psychological biases in deliberation

Given the limitations of the Europolis questionnaires, our study focuses on motivation, cognitive sophistication, and group dynamics. First, motivation is a crucial factor for deliberative performance (e.g., Neblo et al., 2010). High motivation and high involvement is conducive to "central reasoning", leading to a willingness to diligently consider information and arguments; by contrast, low motivation and low involvement is conducive to "peripheral reasoning" and the reliance on information shortcuts (Petty et al., 1983; see also Petty and Cacioppo, 1984). There is, however, some controversy whether high or low involvement increases or decreases deliberative quality. Fung (2003) acknowledges that one possibility is that individuals with low stakes in a discussion (but a basic motivation to engage with the topic under discussion) will be the better deliberators, since low stakes are conducive to dispassionate attitudes and open-mindedness. This view is in line with classic deliberative theory, emphasizing calm and dispassionate reasoning. But Fung (2003) rather supports the opposite view: participants with high stakes may "invest more of their psychic energy and resources into the process and so make it more thorough and creative" (p. 345). We proxy motivation with political interest and salience.[5] While political interest touches upon a more basic motivation to engage in discussion, salience is closely related to involvement and the personal relevance of the issue under discussion. Second, participants with a higher level of prior knowledge on the issue at hand may have a broader argumentative repertoire which may positively influence their deliberative behavior (Cappella et al., 2002; see also Hansen, 2010). Third, group composition and resulting dynamics might also play a role (see, e.g., Karpowitz et al., 2012). Despite the fact that deliberative polls are construed in a way to yield a uniform treatment onto participants, group composition and facilitation style might still vary in practice. In our study, we will take advantage of such variation and explore the effects of more and less heterogeneous groups as well as less vs. more dominant facilitation styles on deliberative quality. We expect that more heterogeneous groups have a higher quality of deliberation. Following epistemic democrats, a higher diversity of viewpoints should lead to a more dedicated exchange of views (see, e.g., Mercier and Landemore, 2012). Facilitation, in turn, is an understudied and under-theorized topic (see Moore, 2012). According to Mansbridge et al. (2006: 30–32), the primary goal of facilitation should be non-domination. Indeed, a dominant facilitator may free participants from engaging in discussion, leading to

[5] This is not an exclusive account of motivation. Other factors such as conflict avoidance or the need for cognition are also important antecedents of motivation (e.g, Neblo et al., 2010). Yet the Europolis questionnaire did not include measures that encompass these psychological constructs. At this point, we restrict the theoretical discussion to the available measures to capture motivation.

what Ryfe (2006: 88) has called "scattershot discussions" where participants tend to speak shorter and are stronger oriented towards the facilitator than towards other participants.

2.3 Measuring Deliberative Quality

While there have been some systematic and in-depth explorations of the deliberative process in legislatures (see, e.g., Esterling, 2011; Mucciaroni and Quirk, 2006; Steiner et al., 2004), there is a dearth of similar research in the context of citizen deliberation (for exceptions, see Dutwin, 2003; Schneiderhan and Khan, 2008; Siu, 2009). In recent years, however, it is becoming increasingly standard to ask participants of citizen deliberation how they evaluate the quality of the deliberative process (see Andersen and Hansen, 2007; Grönlund et al., 2010; Ratner, 2008; Sanders, 2012). In general, evaluations of such deliberative events tend to be positive with participants rating the quality of deliberation as high. Focusing on the Danish poll on the Euro, Andersen and Hansen (2007: 542) find a "high degree of quality of procedure": participants reported that the discussion was focused on general principles and the common good, even though self-interest was not absent in the discussions. We acknowledge that self-evaluation is an important pathway to evaluate the quality of the deliberative process (see Bevir and Ansari, 2012; Neblo, 2007b). As King (2009: 6) argues, Habermasian discourse ethics "is interested in the intersubjective achievement of understanding and process of decision-making". According to King (2009: 6), a proper evaluation of deliberative quality must assess "how this is perceived by other participants – not merely the subjective speculations of outside observers". Yet, self-evaluation of deliberative quality can suffer from serious methodological flaws: first, self-evaluations may not be reliable, particularly given that participants know what the practitioners want to hear (Muhlberger, 2007). Second, the participant perspective may also be insufficient from a philosophical perspective. Participants may think that the quality of reasoning was good, whereas philosophers would judge the respective reasoning as insufficient by their own theoretical standards. Since classic deliberation's theoretical focus is on the realization of philosophical ideals, we think that it is imperative to complement the participants' perspective with an external and philosophically grounded evaluation analysis of deliberative quality. A pioneering effort in this regard has been made by Siu (2009). Drawing from the Discourse Quality Index (DQI), Siu analyzed the transcripts of several online and face-to-face deliberative polls in the United States. Yet, Siu's analysis has several limitations: it mainly focuses on justification rationality, while neglecting other deliberative elements such as respect or common good orientation. Siu also omitted any statement about the abilities of

citizens to reach deliberative standards in the discussions she analyzed. The goal of this article is to provide a more comprehensive assessment of deliberative quality.

2.3.1 Discourse Quality Index 2

In this study, we will assess deliberative quality on the basis of the Discourse Quality Index (DQI; Steenbergen et al., 2003). The DQI allows for a quantitative content analysis at the level of individual speeches of recorded discussions, measuring the quality of deliberation with six indicators that are anchored in Habermasian discourse ethics. The DQI captures deliberation via justification rationality, common good orientation, respect toward groups, demands and counterarguments, and constructive politics. Not only has the DQI met with considerable support from deliberative theorists (Habermas, 2005; Thompson, 2008), it is also the most encompassing and most widely used measure of deliberation (see, e.g., Bächtiger, 2005; Caluwaerts, 2012; Lord and Tamvaki, 2013; Steiner, 2012).

However, the original DQI has been developed for the analysis of parliamentary debates and is strongly rooted in a classic and Habermasian-inspired understanding of deliberation. This dual focus on the political sphere and on classic and Habermasian-inspired deliberation raises questions of how well the DQI can be applied to citizen deliberation. Yet, we do not see any need to abandon the classic standards. Rather, "standards must be loosened to accommodate the more informal character of the nongovernmental parts of the deliberative system, but in this loosening, they do not lose their character" (Mansbridge, 1999: 213). As our empirical analysis will demonstrate below, fears that the DQI sets overly demanding standards that no citizen deliberator can achieve are largely unsubstantiated. We shall make some adaptations in the evaluation procedure to take into account the constraints of citizen deliberation: we do not expect that citizen deliberators achieve the (demanding) standards of classic and Habermasian-inspired deliberation all the time or on average, but only once during an entire discussion (see below). By the same token, the DQI might also profit from being enriched with alternative forms of communication. Feminists and difference democrats have argued that rational discourse may put overly demanding requirements on ordinary citizens. Many deliberationists nowadays consider "alternative" communication modes such as story-telling or testimony as valid and even desirable deliberative practices for people with marginalized points of view (e.g., Sanders, 1997; Young, 2002; see Polletta and Lee, 2006). We think that such developments in deliberative theory and practice must be reflected in an evaluation of citizens' deliberative capacities as well and shall thus add one element of expanded notions of

deliberation, namely "story-telling". In the following, we briefly describe the various components of the updated *Discourse Quality Index* (DQI 2).

(1) *Justification rationality*: In classic deliberation, high justification rationality is a key criterion for high deliberative quality. Since the "ideal speech situation" itself has no content, one cannot apply external standards to what constitutes a good reason. Hence, this standard focuses only on the syntactic structure of argument and judges to what extent a speaker gives complete justifications and thus makes his speech accessible to rational critique. It distinguishes between five levels of justification rationality: (0) no justification; (1) inferior justification where the linkage between reasons and conclusion is tenuous (this code also applies if a conclusion is merely supported with illustrations); (2) qualified justification where one complete justification is offered and a linkage between reasons and conclusion is made; (3) sophisticated justifications where a problem is examined from different viewpoints or in-depth.

(2) *Common good orientation*. In the eyes of many deliberative democrats, good reasons should entail a common good orientation. The original DQI measures whether arguments are cast in terms of narrow group or constituency interests, whether there is neutral reference or mixed reference (i.e., reference to both narrow group interest and common good), or whether there is a reference to the common good. With regard to the common good, the original DQI focuses both on the common good stated in utilitarian terms, i.e., as the best solution for the greatest number of people and the common good expressed through the difference principle, i.e., the common good is served when the least advantaged in a society are helped (Rawls, 1971). In the context of a transnational discussion such as Europolis, the categories need to be refined, however. We distinguish among references to country interests (1), references to two sorts of interests, country and European interests (1.5), European Union (2) and world community interests (3), and the absence of such references (0).

(3) *Respect and interactivity*. Another key component of deliberative quality is respect. Macedo (1999: 10) regards the recognition of the "merit in [the] opponents' claims" as being one of the principal purposes of deliberation. The original DQI measures respect with three dimensions: respect toward groups, demands, and counterarguments. This three-pronged conceptualization of respect has met with some difficulties in the practice of coding. For instance, the distinction between respect toward demands and respect toward counterarguments is not always so clear. 'Economies of speech' may lead actors to focus on either demands or counterarguments. In the context of citizen deliberation, we simplify the

task by focusing merely on respect toward other participants' arguments. Classic deliberative theory is about arguments rather than demands and claims (which belong more to the realm of bargaining). By focusing on respect toward other participants' arguments, we also link respect to interactivity, a hitherto neglected aspect in the empirical measurement of deliberation. Interactivity means that people respond to what others have said (see Goodin, 2000: 91). Specifically, we measure whether speakers include other participants' arguments but degrade them (0), whether speakers ignore other participants' arguments (1), whether they include other arguments in a neutral fashion (2), and whether they value other participants' arguments (3). In addition, we also focus on respect toward groups, which is a reflection of Habermas' (1991: 73) emphasis on empathy and solidarity. Respect toward groups implies that participants, either implicitly or explicitly, acknowledge the needs and rights of different social groups. The DQI captures whether a speaker degrades other groups (0), neither degrades nor acknowledges them (1), whether she is neutral (2) or whether she shows explicit respect toward them (3). In the context of the Europolis discussions on immigration, respect toward groups concerned third country migrants.

(5) *Questioning*: Deliberative democrats also emphasize the importance of inquisitiveness (see Cohen, 1989). The original DQI has not considered inquisitiveness as an essential aspect of an interactive deliberative process. We rectify this deficiency, whereby we operationalize inquisitiveness via questioning. Questioning has an informational and a critical function (see Ikuenobe, 2001: 334), even though the two frequently complement each other. We code whether a speech contains an informational or critical question (coded as 1) or not (coded as 0).

(6) *"Story-telling"*: As mentioned before, difference democrats and feminists have stressed the need to admit wider forms of communication – such as story-telling, testimony, or rhetoric. According to Polletta und Lee (2006), story-telling is the most important component of alternative forms of communication. In order to capture story-telling empirically, we measure whether participants use personal narratives or experiences (see Stromer-Galley, 2007).

For the construction of our measure, we leave three components of deliberative quality aside: participation equality, constructive politics, and truthfulness (or, sincerity). First, even though participation equality is a crucial deliberative standard, it does not really capture the formal

quality of arguments (see Estlund, 2000).[6] In addition, interpersonal deliberation requires a conception of equality as "partner in dialogue" (Warren, 2001: 81). In other words, whether individuals are able to participate at equal rates does not exclusively depend on their deliberative abilities but as well on others' willingness to consider their contributions on equal merits. Nonetheless, we will make a first attempt and relate our measure of quality to measures of participation equality, namely the frequency and length of speaking (see section 2.6). However, for a full account on participation equality in Europolis, see Chapter 3. Second, since deliberative opinion polls do not entail decision-making at the end, we expect that constructive politics, i.e. the pursuit of a rationally motivated consensus, will only play a negligible role in such discussions. For this reason, we have excluded this component from the analysis. Finally, Habermasian deliberation would also require truthfulness or sincerity, which is the absence of deception in expressing intentions. However, to judge whether a speech act is truthful is to make a judgment about a person's true versus stated preferences. This is exceedingly difficult, since true preferences are not directly observable. The speculative nature of such a judgment is bound to introduce large amounts of (possibly systematic) measurement error. Moreover, in the context of deliberative polls where participants do not make consequential decisions the truthfulness problématique is considerably reduced. Therefore, we exclude the truthfulness criterion from our analysis as well.

2.3.2 Setting Thresholds for High and Low Deliberative Quality

Recent years have witnessed an increasing demand for setting one or more "threshold" or "cut" values for deliberation (see Black et al., 2010). Regarding the DQI, Drzyek (2007: 244) has noted: "in applying the discourse quality index, it is hard to say whether the deliberation in any of the cases analyzed is actually good enough by any theoretical standards. The index is just a comparative measure". In this article, we address this deficit and define thresholds for the various DQI indicators at the level of individuals. Although this will put us in a position to judge whether individuals were able to reach high standards of deliberative quality, we cannot infer whether the discussion as such displayed a sufficient level of deliberative quality.[7]

[6] The original DQI captures equal participation by asking whether speakers were able to utter their claims freely or whether they were interrupted by other discussants (Steenbergen et al., 2003). Although this might be a suitable measure to capture participation equality in the realm of legislatures, it does not work in (simultaneously translated) deliberative polls where interrupting another participant is technically impossible.

[7] See also Appendix C for summary statistics on the overall discourse quality.

The threshold level problem is intertwined with a level of analysis problem. To date, the quality of deliberation was only checked at the level of individual speeches. To achieve an overall maximum score, every speaker would not only have to extensively justify her demands and arguments in every single speech, she would also have to be simultaneously oriented toward the common good and be respectful at all times. Even staunch advocates of deliberation might agree that this is a conceptual impossibility, ignoring 'economies of speech' and the fact that in good conversations, arguments are not repeated all the time. Moreover, it would be overly demanding to expect that ordinary citizens constantly reach the various deliberative standards during discussion. Therefore, we apply a *holistic approach* which analyzes the maximum deliberative performance of a speaker in an *entire discussion*. In other words, we do not expect that citizens live up to classic deliberative standards all the time or on average, but only that they achieve the various quality standards *at least once* in the discussion.

To identify high and low quality standards for the DQI indicators, we draw from a classic and Habermasian-inspired conception of (maximum) deliberative quality. With regard to classic deliberation, there is broad agreement in the literature that this type of communication entails complex reasoning and is geared towards finding common understanding and common values (see Bächtiger et al., 2010a; Mansbridge et al., 2010). Translated to the DQI indicators, *classic deliberation means that participants offer sophisticated rationales, refer to the common good, show explicit respect toward other participants' arguments as well as to other groups, as well as question what others have claimed*. Concretely, all DQI indicators are dichotomized in accordance with these cut levels. Coding examples and the dichotomization procedure are detailed in Appendix A. Moreover, the expectation in classic deliberation is that the different quality standards occur *simultaneously*, i.e., good deliberators should ideally comply with all quality standards. There is also *equivalence* among the different standards, i.e., no priority or differential weight is given to specific indicators of deliberative quality. The rationale behind simultaneity and equivalence is the following: if a speaker only gives reasons but is never respectful (or, even does this in a disrespectful way) then this would be indicative of a rhetorical mode of communication rather than reflexive rational discourse. Or, if a speaker is only respectful but is not able to provide complex reasons for his position, then this might be a sign of deference rather than autonomous reflection. Finally, it is an open question whether classic deliberation can also involve story-telling so that positions and arguments are conveyed in a comprehensible fashion. As Neblo (2007b) convincingly argues, there is an opening in Habermasian theory towards alternative forms of communication if the

latter are necessary to convey a specific message or to equalize differences in cognitive abilities. The inclusion of story-telling in a measure of classic deliberation can give rise to two scenarios: if story-telling is aligned with the other components of classic deliberative quality, then it is a sort of a "rhetorical" device that higher skilled deliberators employ as well in order to convey their messages in a comprehensible fashion. By contrast, if story-telling is not aligned with the other components of classic deliberative quality, then it might represent a distinct form of expression that is primarily used by lower skilled deliberators or by specific social and cultural groups (as it was originally imagined by feminist critics of deliberation; see Sanders, 1997; Young, 2002).

2.3.3 Item Response Analysis

In order to identify deliberative quality empirically, we first check whether citizens meet the pre-defined standards of classic and Habermasian-inspired deliberation and whether the different standards actually form a compound phenomenon. Concretely, we explore whether the different indicators of the DQI represent a reflection of a latent variable of deliberative quality ranging from high to low deliberative quality. In order to construct a latent variable of deliberative quality, we use Bayesian item response analysis. Item response analysis was originally developed in psychology and educational science to measure latent psychological constructs based on responses to different question items (Rasch, 1980). A typical latent construct is intelligence. Denoting the intelligence level of respondent *i* by β_i and the difficulty level of item *j* by α_j, the probability that *i* gives a correct answer to *j* can be modelled as follows:

$$\text{Logit}(\text{Prob.}(i \text{ gives the correct answer to } j)) = \gamma_j (\beta_i - \alpha_j)$$

This equation assumes that the probability of a correct answer is given by the extent to which the degree of intelligence exceeds the difficulty of the question. Hence, the larger the difference between the difficulty of the question α_j and the degree of intelligence β_i, the higher the probability that a correct answer is given by respondent *i*. In this context, γ_j is often called the discrimination parameter since it represents the impact of the latent dimension on the response. If $\gamma_j = 0$, there is no relationship between the identified latent dimension and the response category.

This logic can be profitably applied to our research purpose. We can interpret deliberative ability similar to intelligence levels in educational science, namely how well citizens are able to achieve the various standards of deliberative quality (justification rationality, common good orientation, respect, etc.). This also bears the advantage that we do not expect that *all*

participants must achieve the various quality standards in one single speech. Rather, we can make the more realistic assumption that some components of deliberative quality may be more difficult to achieve than others. Item response models take into account that different individuals can exhibit different relations with the latent construct (Reise et al., 2005). To demonstrate that deliberative quality is a latent construct, the item response functions of the various deliberative standards should display similar slopes for the discrimination parameters.

Compared to other aggregation methods, such as additive indices and factor analytic methods, the item response model has several distinctive advantages. Compared to an additive index, the item response model does not a priori assume that all deliberative standards are identically relevant for deliberative quality. Even though philosophers tend to invoke simultaneity and equivalence among the different deliberative standards, this is an assumption that needs to be tested rather than assumed. Compared to factor analytic methods, which also relate deliberative standards to the underlying dimension via differentiated loadings, the item response function provides a more intuitive understanding of this relationship, especially by yielding estimates of both difficulty and discriminating power. While conventional factor analysis only models the covariance of the items, item response theory models the various items directly as a function of the latent construct (Jackman, 2001; Treier and Jackman, 2008). By considering thresholds (and yielding a difficulty parameter for each threshold), item response analysis is in an excellent position to test for philosophical constructs such as deliberative quality, which entail the idea of thresholds in the form of "high" and "low" quality. Finally, factor analytic methods are not appropriate for our dummy coded data since normally distributed error terms are assumed.[8]

It is well known that item response analysis suffers from an identification problem (Jackman, 2001). To address this problem, we apply a Bayesian framework and use prior information. Since we lack guidance for the construction of strongly informed priors, we adopt an empirical approach. Concretely, we set a truncated normal distribution for the discrimination parameter in order to force the corresponding parameter on a negative or positive side of the scale. The choice of the items whose discrimination parameters are set with an informative prior was conducted in a stepwise fashion until the identification problem is completely solved.

[8] One important assumption of item response models is the local independence of items (e.g., Treier and Jackman, 2008). Aggregating the speech act level data to the individual level (as demonstrated in Appendix A) makes our DQI indicators independent from each other: theoretically, any deliberative standard can be met irrespective of one's success in reaching another standard. Empirically, a correlation analysis of residuals shows that our data do not violate the local independence assumption.

2.4 Data

2.4.1 Research Setting: Europolis

We analyze the potential of citizen deliberation on the basis of *Europolis*, a pan-European deliberative poll which took place in Brussels in May 2009. During the three-day event, participants from all 27 EU-member states discussed the topics of third country migration and climate change. Europolis employed the well-known method of deliberative polling (DP) (see, e.g., Fishkin and Luskin, 2005; Fishkin, 2009). DP incorporates several institutional devices designed to support citizens' deliberative potential: participants receive balanced briefing material and can question experts and politicians, which may help to enhance their epistemological resources. In addition, facilitators guide the small group discussions in order to establish an open and considerate atmosphere for discussion.

The Europolis project was administered in the following way (Isernia et al., 2013): it started with interviewing a random sample of about 4'300 European citizens in April 2009. About 1'300 of them were randomly selected to be part of the control group. The remaining 3'000 were asked at the end of the interview whether they show interest in participating at a deliberative polling event. Out of the ones showing interest, a random sub-sample of 400 citizens, proportionally stratified according to the number of seats per member state in the EU Parliament, was invited to participate at the DP event. 348 out of them actually took part. During the three-day event in Brussels, the participants were randomly assigned to 25 small groups for discussion. The discussions were simultaneously translated in the languages spoken in the respective groups. Due to limited translation capacities, a maximum of three different languages was allowed per discussion group. This limited the random assignment to small groups to some extent and as a consequence, group discussions varied considerably in degree of transnational heterogeneity.

Since analyzing deliberative processes empirically is a highly demanding and time-consuming affair that required that we first transcribe the audio discussions, we refrained from analyzing all 25 small group discussions. Rather, we took a purposive sample and limited our analysis to the general discussion on immigration and the more specific discussion on policy options that address third country migration.[9] We decided to focus on discussions between citizens of new EU member states (2004-) and citizens of the older states of the EU. Moreover, we also

[9] In doing so, we leave aside two short sessions where participants were asked to formulate questions to experts and politicians, as well as a session where participants were encouraged to study and discuss a variety of European party manifestos on third country immigration.

wanted to focus on discussions between citizens from Western European and Southern European countries. By simultaneously holding the variation within the group of Western European participants as small as possible, the group selection enables us to test for cultural and "newcomer" biases in deliberative behavior.[10] Although our subsample is limited in national diversity compared to the full sample of participants, it is largely comparable to the rest of the Europolis groups with regard to socio-economic and psychological variables. Appendix D displays properties of the groups that have been analyzed and compares them to the rest of the sample.

2.4.2 DQI Indicators and Latent Deliberative Quality

In a first step, we coded every speech act that participants delivered in the main discussion sessions on immigration by relying on the DQI coding scheme (see section 2.3.1 and Appendix A). An inter-coder reliability test on the various DQI standards was conducted prior to coding the full database. Three coders independently evaluated 80 speeches. As we detail in Appendix B, reliability proved to be good to excellent, i.e., there is generally broad agreement where a particular speech act falls on the six indicators. Cohen's kappa, which controls for inter-coder agreements by chance, ranges from 0.70 for respect toward other participants' arguments to 0.85 for story-telling. Notice, however, that for two indicators, level and content of justification, a first reliability test yielded only satisfactory results. Consequently, we have clarified the disagreements, re-coded the data, and then performed a second inter-coder reliability test for level and content of justification. This test then yielded a very good reliability for the two indicators. In sum, the high inter-coder reliability rates greatly increase the confidence one can place in the subsequent transformation and analyses performed with the DQI indicators. After the coding, we aggregated the speech act data to the level of individuals and extracted participants' "best performance" on every standard of deliberative quality. By dichotomizing the individual level data as detailed in Appendix A, we created variables that reveal whether individuals were able to meet the previously specified thresholds of high deliberative quality. This was the input data for the Bayesian Item Response Analysis (section 2.3.3). Via Bayesian Item Response Analysis, we obtained ideal points of a latent construct of classic and Habermasian-inspired deliberative quality. The ideal

[10] The participants from Western Europe mainly originate from one of the founding states of the European Union. We excluded groups including participants from the United Kingdom, since this would introduce an additional dimension of variance in relation to EU history and experiences with migration (see Freeman, 1995; Geddes, 2003).

points serve as dependent variable for the bias analysis (see next subsection). Details of the item response analysis are presented in the results section.

2.4.3 Predictor Variables

In order to explore whether deliberative abilities differ among socio-economic, cultural or psychological variables, we conduct multilevel regression analysis. The multilevel analyses take into account that participants in Europolis are nested in discussion groups and that two predictors, group composition and facilitator activity, are located at the group level. Given our dependent variable of latent deliberative quality (see above), we estimate hierarchical linear models.

Regarding predictor variables, we focus on gender (1=female; 0=male), age (measured in 10 years), education (measured as the age at end of the education process), working-class (1=working-class; 0=other), political interest (measured on an 11-point scale from 0 ("not at all") to 10 ("passionately"), salience (seriousness of the immigration problem, measured on a scale from 0 ("no problem at all") to 10 ("the most serious problem we face")), and prior knowledge (number of correct answers to three knowledge questions on immigration).[11] A variable for briefing material captures whether participants have read the information material. The scale ranges from 1 ("have not read or only glanced at it") to 5 ("have read all or nearly all of it"). A categorical variable allows us to distinguish between participants coming from Central Eastern, Southern or Western European countries. With regard to group composition and group dynamics, we focus on group heterogeneity and facilitator style. To obtain a measure of group heterogeneity for our mixed national groups, we first evaluated group heterogeneity on the basis of the *Ethnolinguistic Fractionalization Index* (ELF), by measuring the share of national groups out of the total number of national groups that were present in the small discussion groups.[12] The ELF ranges between zero and one, with higher values indicating higher levels of fractionalization and diversity. Yet the ELF indicator does not take into account another crucial feature of European diversity, namely the broader cultural background of participants and the possibility that experiences with immigration types and policies differ across countries or regions (see Freeman, 1995; Geddes, 2003). Therefore, we weight the group scores obtained on the basis of the ELF indicator as follows:

[11] The three knowledge questions in the Europolis questionnaire concerned the definition of a Blue card worker, the current form of the EU immigration policy and some figures on the EU's immigration population.

[12] Ethnolinguistic Fractionalization Index (ELF)= $1-\Sigma s_i^2$ where s_i is the share of group i out of a total of n groups. The ELF measures the probability that the two randomly selected individuals from the entire participant pool will be from different groups.

groups consisting of Western, Central Eastern, *and* Southern Europeans obtain a weight of 3; groups consisting of Western and Central Eastern Europeans or Western and Southern Europeans obtain a weight of 2; and groups where all participants originate from the same region obtain a weight of 1. In so doing, groups with a high diversity of cultural backgrounds receive a higher score than groups with low diversity. Facilitator performance is captured by the group's average opinion on the basis of the following question "My small group moderator sometimes tried to influence the group with her / his ideas". Participants could evaluate facilitator performance on an 11-point scale where 0 indicates complete disagreement and 10 complete agreement to the statement.

2.5 Results

Before we present the results, let us first sketch an optimistic and a pessimistic scenario for deliberative abilities of ordinary citizens. In the optimistic scenario, the different deliberative standards – justification rationality, respect etc. – form a compound phenomenon and genuine deliberative personalities scoring high on all standards exist. High quality deliberation is not constrained to a small number of citizens, and especially socio-economic and cultural factors – which are less changeable than psychological characteristics - are not or only weakly correlated with deliberative abilities. In the pessimistic scenario, deliberative quality is not a compound phenomenon, but falls apart into its various components (justification rationality, respect etc.). Alternatively, deliberative quality forms a compound phenomenon, but high deliberative skills are constrained to a very small circle of citizens and are strongly correlated with socio-economic and cultural characteristics. From a statistical vantage point, we should see statistical models with little statistical and substantial significance in the optimistic scenario, and statistical models with high statistical and substantial significance in the pessimistic scenario (for a similar research strategy to uncover biases in deliberation, see Luskin et al., 2002: 478–79). Optimistic or pessimistic scenarios for deliberative theory also hinge upon the question whether biases in deliberative abilities are amplified by other biases, especially participation equality, and transfer to the outcomes of deliberation as well. In short: if biases in deliberative ability combine with a lower "standing" in the process and also lead to differential outcomes between high and low skilled deliberators in terms of openness towards change, then pessimism about deliberation's unbiased functioning is amplified. We shall return to these points in section 2.6.

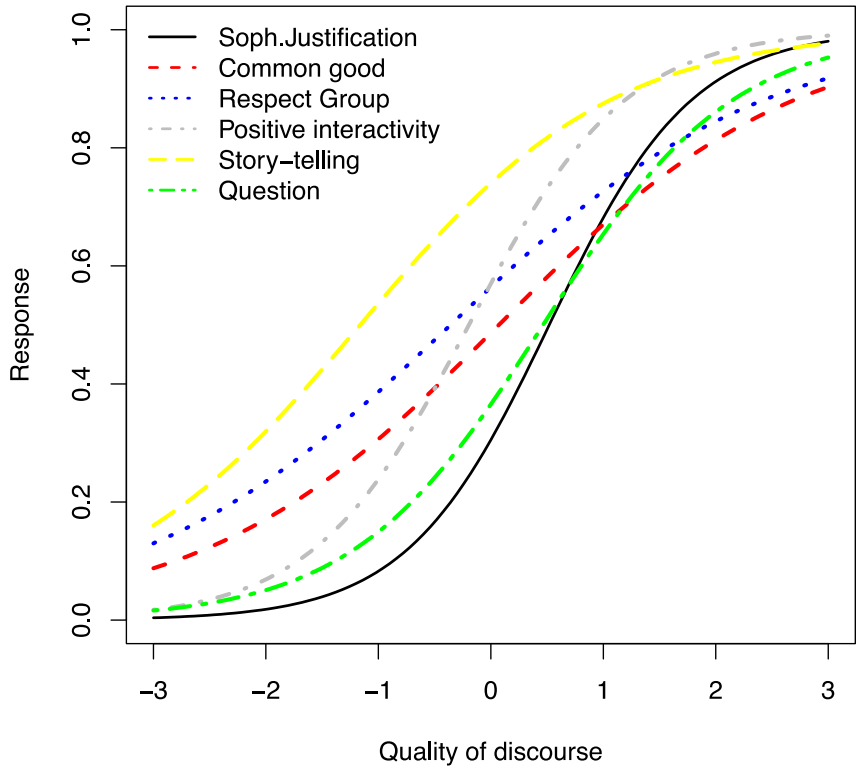
2.5.1 Foundations of the Genuine Deliberator

We first explore whether a genuine citizen deliberator exists in the real world. On the basis of a *Bayesian Item Response Analysis*, we check whether the different DQI components – justification rationality, respect etc. – form a latent dimension of deliberative quality.[13] In order to achieve identification of the model, we used a standard normal distribution censored in non-zero values as a prior for the discrimination parameters. Due to this specification, a higher score of an ideal point corresponds to a higher level of deliberative quality. We ran three Markov chains with 7000 iterations from different initial values, respectively. We discarded the first 2000 iterations of each chain as burn-in and assessed convergence by visual inspection as well as by using the improved Brooks-Gelman-Rubin convergence diagnostics. We did not detect any sign of non-convergence (results available upon request).

Figure 2.1 presents the item response functions relating the latent dimension to the response probability of each indicator of deliberative quality. We see that all item response functions have similar positive slopes for the discrimination parameters, suggesting that the indicators form a latent construct. Put differently, participants achieving the more difficult items (sophisticated justification) also have a higher probability to achieve the easier ones (e.g., respect toward other participants' arguments). This means that against skepticism from personality psychology, a genuine deliberator scoring high on justification rationality, common good orientation, respect, empathy, and inquisitiveness exists in reality. The relatively steep discrimination parameters also indicate that participants with lower deliberative ability differ considerably from participants with higher deliberative ability. A closer look at the difficulty parameters indicates that the demanding standards of classic deliberation were far from being rare events in the Europolis discussions. Zero on the horizontal axis corresponds to the percentage of participants reaching this standard. We see that story-telling was the easiest item, with a 70 percent chance that Europolis participants reached this standard. This is in line with Ryfe's (2006) study of five National Issues Forums, finding that the primary form of reasoning is story-telling. For common good orientation, explicit respect toward other participants' arguments, and explicit respect toward third country migrants, there is about a 50 percent chance that participants reached these standards. Sophisticated justification was the most difficult item, with a 37 percent chance that participants reached this standard.

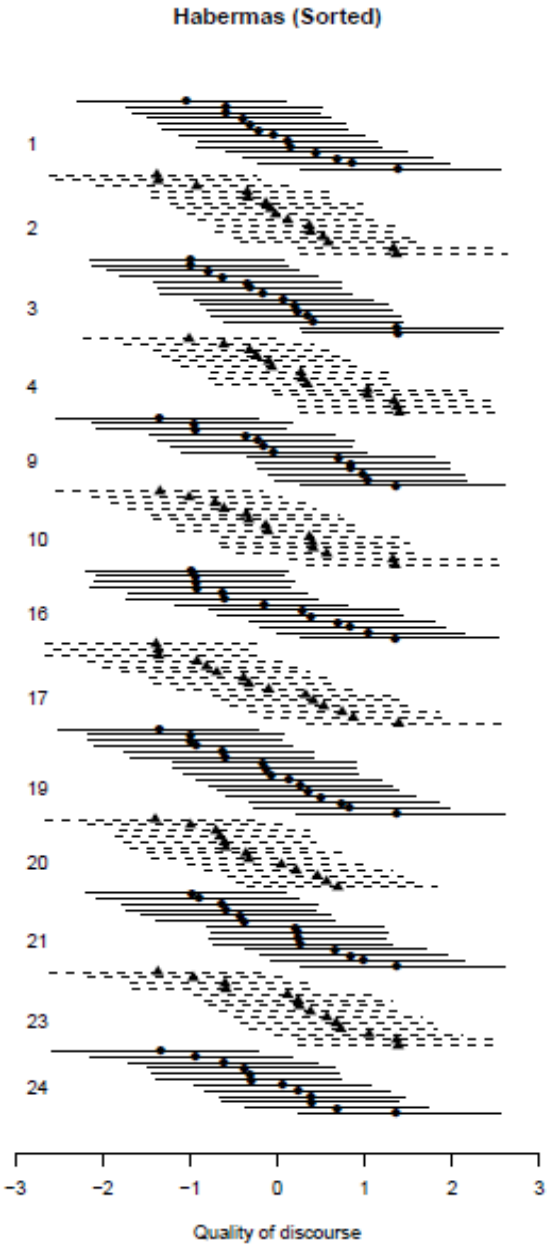
[13] The Bayesian Item Response analysis was performed in WinBUGS, a free software package for Bayesian estimation.

Figure 2.1: Response Function for each Item



An intriguing result is that story-telling - even though it represents the easiest standard - has a strong relationship with the latent dimension. This means that story-telling is a partial complement to justification rationality, i.e., people who make sophisticated justifications also use story-telling. This provides a hint that the classic distinction between rational discourse and alternative forms of communication may be misled, since high-skilled deliberators also use personal experiences to back their positions and arguments. Of course, the chance rates only display whether the diverse standards of classic deliberation were reached once during an entire discussion. But the results are still quite striking. Theoretically, one might have expected that classic deliberative standards all form difficult items of an intelligence test. But this is clearly not the case, and the standards of classic deliberation are far from being utopian standards that no citizen deliberator can achieve. Surely, the number of genuine high quality deliberators is quite low: focusing on the raw data, the number of participants reaching all six standards is only 10 percent.

Figure 2.2: Individual Latent Deliberative Quality (Sorted by Group)



In Figure 2.2 we explore whether high and low deliberative quality varies according individuals and discussion groups. Notice that scores on the latent variable cannot be interpreted in an absolute way; but the more we move to the right-hand side of the continuum, the higher the deliberative quality (and vice versa). Individual points with 90% credible intervals correspond to individual participants nested in 13 discussion groups (we use the Europolis group classifications to denote the various groups).[14] In accordance with the Bayesian approach, we should not interpret the overlapping credible intervals as the absence of a statistically significant difference, but the magnitude of overlap as the probability of indifference of ideal points. In other words, the less overlap the higher the probability that

[14] Credible intervals are the Bayesian analogue to confidence intervals in conventional frequentistic statistics.

participants differ in their ability to deliberate. Participants are further sorted based on the average deliberative quality in each group. The panel clearly displays that every discussion group involves participants with high and low deliberative ability. By the same token, there are no outstanding differences in deliberative quality among the 13 discussion groups when comparing maximum performances of individuals.

2.5.2 Exploring Biases

In order to explore whether particular socio-economic, cultural or psychological variables predict individual deliberative ability, multilevel models were estimated for our latent construct of deliberative quality. As the random intercept model in Table 2.1 displays (Model 1), there is only one statistically reliable effect: working-class participants score lower on deliberative quality than participants from higher classes ($p<.05$). This confirms sociologists arguing that working-class people possess a *repertoire* of speaking and arguing which may not so easily align with classic and Habermasian-inspired forms of deliberation. We have also probed for interaction effects among socio-economic and cultural variables, and found two. Working-class participants from both Central Eastern Europe and Southern Europe (Spanish and Portuguese) do not reach the same levels of deliberative quality than other participants. Both interactive effects are substantively large and statistically significant ($p<.05$). This means that clearly less privileged people in the European "polis" - lower class participants from the European periphery – were also the least skilled deliberators. Indeed, when focusing on the raw data, we see that working-class participants from CEE, for instance, did not provide a single sophisticated rationale in the entire discussion. From a bias perspective, this is a worrisome finding, meaning that already disadvantaged people have trouble to adapt to deliberative modes of interaction, which turns citizen deliberation into a fragmented exercise.

Notice, however, that Central Eastern and Southern Europeans (Spanish and Portuguese) *per se* do not perform worse than Western Europeans. Region turns into a significant predictor only when it is interacted with class (Model 2). Thus, at a general level, popular claims that culture is a powerful predictor of deliberative quality are not corroborated.

Table 2.1: Antecedents of Deliberative Quality

	Model 1	*Model 2*
Fixed effects		
Individual variables		
Intercept	0.295 (0.499)	0.351 (0.486)
Men	0.121 (0.131)	0.181 (0.129)
Education	0.008 (0.014)	0.005 (0.014)
Working-class	-0.381 (0.158)*	0.082 (0.216)
Age (in 10 years)	-0.068 (0.039)$^{+}$	-0.061 (0.039)
Central Eastern Europeans (CEE)	-0.192 (0.176)	-0.022 (0.196)
Southern Europeans (SE)	-0.203 (0.186)	0.029 (0.198)
Working-class x SE		-0.744 (0.356)*
Working-class x CEE		-1.108 (0.373)*
Salience	-0.049 (0.027)$^{+}$	-0.052 (0.026)*
Knowledge	0.085 (0.070)	0.111 (0.069)
Interest in politics	0.055 (0.032)$^{+}$	0.051 (0.031)$^{+}$
Briefing material	0.000 (0.055)	-0.009 (0.054)
Group level variables		
Group heterogeneity	-0.026 (0.149)	-0.042 (0.145)
Facilitator performance	-0.034 (0.080)	-0.046 (0.078)
Random effects		
Individual level	0.579 (0.761)	0.549 (0.741)
Group level	0.000 (0.000)	0.000 (0.000)
AIC	427.9	422.5
Log Likelihood	-198.9	-194.3

Note: Multilevel linear models estimated with the lme4 package in R (Bates et al. 2011). Estimated standard errors are shown in parentheses. $^{+}$p<0.1, *p<0.05, **p<0.01. Individual N=159; Group N=13.

Nonetheless, one may still wonder whether the two interaction effects are the product of different "speech cultures" (with working-class people from Central Eastern and Southern Europe having a different way of speaking and arguing than other participants), or whether they represent a "newcomer" effect, especially mirroring the inexperience of Central Eastern Europeans with pan-European affairs (see Naurin, 2010). Of course, no definitive answer can be given here. But the fact that the substantive effect is clearly stronger for Central Eastern Europeans than for Southern Europeans which have a longer history of EU membership than Central Eastern Europeans provides a first hint that the effect may be more due to newcomer status than due to different "speech cultures". Next, we find that higher salience levels lead to lower deliberative quality. This effect is marginally significant in Model 1 (p<0.10) and significant in Model 2 (p<0.05). This is an intriguing result, supporting classic deliberation's claim that it is dispassionate attitudes – and not passionate ones – which are conducive to

higher deliberative quality. Furthermore, and in accordance with our expectations, we find that higher political interest (a proxy for general motivation to engage with the topic under discussion) is slightly associated with higher deliberative quality. Contrary to our expectations, lower age is associated with higher deliberative quality. However, these two effects are only marginally significant ($p<0.10$). No statistically reliable effects are detected for gender, education, prior knowledge and the intensity of studying the briefing material. Given its (enduring) prominence in the literature (see Karpowitz et al., 2012), the absence of gender effects is an important result. It defies feminists' claims and empirical findings in psychological experiments that women are disadvantaged in deliberative processes. In case of education, the absence of any effect seems surprising at first glance. However, in the Europolis case, we suspect that this results may be due to the particular way the Europolis questionnaire measures education – namely as "years of education". This operationalization does not properly disentangle participants having higher education and a university degree from those who have not; this distinction may be the driving factor behind high and low deliberative quality, however.[15] The absence of information effects (prior knowledge and intensity of studying the briefing material) is noteworthy too: it means that biases in deliberative abilities cannot be easily compensated by enhancing the knowledge of participants. Finally, at the level of groups, neither facilitator activity nor group heterogeneity affects the quality of deliberation. Eventually, this result may not be so surprising. As we saw before, variation across the thirteen discussion groups is very small, a fact that is also documented by the extremely small variation at the group level.[16] Moreover, we should not forget that despite variation of group heterogeneity and (perceived) facilitator activity, the comparison here is one of differences-in-degree rather than of differences-in-kind. On the one hand, the random assignment of participants to discussion groups avoids homophily effects (whereby like-minded and otherwise similar people self-select in discussion groups); on the other hand, despite some variation in the application, facilitation styles were still in basic accordance with a general facilitator script.

[15] In accordance with this interpretation, working-class has no statistically significant association with this measure of education (Pearson's $r=-0.18$, $p=0.27$). If the education measure had properly separated between people with higher education / university degree and others, then we should see a much stronger association with working class.

[16] One might object that variance components are underestimated when multilevel model with a small number of level-2 units are estimated via maximum likelihood procedures. However, an ANOVA-Test confirmed that the variance in latent deliberative quality between Europolis groups is negligible ($p=0.825$). In addition, the lme4 package estimates linear models using restricted maximum likelihood (REML), which mitigates possible biases to a considerable degree (e.g., Steenbergen and Jones, 2002; Browne and Draper, 2006). Therefore, we refrained from using Bayesian inference to estimate the multilevel models presented in Table 2.1.

In sum, citizen deliberation in supportive institutional environments such as deliberative polls is not immune against bias, documented by the important class-region effect. But at a general level, citizen deliberation may work better than its most fervent critics have postulated. Not only is deliberation a more widespread ability than commonly assumed, important socio-economic variables such as gender and age are not – or, only weakly associated - with deliberative quality.

2.6 Extensions and Implications

In this section, we put the findings of our bias analysis in perspective, giving context and further meaning to them. First, we relate our measure of deliberative ability to participation equality, which represents another crucial measure of deliberative quality (and bias; see, e.g., Thompson, 2008). Second, we explore whether biases in deliberative abilities also translate into the outcomes of deliberation. Third, we shall discuss the implications of our findings for transnational and global citizen democracy and the design of deliberative events.

First, let us consider the relationship between our latent construct of deliberative quality and participation equality. Recall that we have excluded participation equality from our measure of deliberative quality since it is not directly concerned with the formal quality of argumentation. Participation equality can be assessed in several ways, ranging from frequency and volume of participation (e.g., Siu, 2009; Stromer-Galley, 2007) to equality of consideration, i.e. whether arguments from specific social and cultural groups received adequate consideration from other groups (Pedrini et al., 2013). We focus here on the standard evaluation measure, frequency of speaking and length of speaking.[17] A correlation analysis shows that both frequency of speaking and the total speaking time (measured in seconds) are moderately correlated with our measure of deliberative quality: for frequency of speaking, Pearson's r is 0.65 ($p<0.001$); for speaking time, the coefficient is 0.66 ($p<0.001$). This means that high-skilled deliberators are also the more dominant participants in discussion. In other words, some participants in Europolis not only lacked "articulateness" (Young, 2002: 38), they simultaneously lacked "standing".

Open-mindedness is a defining feature of highly competent deliberative actors (Gutmann and Thompson, 1996). If only deliberatively low skilled and disadvantaged participants shift their preferences while high skilled and advantaged participants do not change their minds at all,

[17] See Chapter 3 for a more encompassing analysis of participation equality. Notice that we have refrained from inserting frequency of participation or speaking time as predictors in our models since these variables are strongly endogenous with our measure of deliberative quality.

deliberative ability would represent a mere "etiquette" (Baccaro et al., 2011). A full-fledged analysis of the outcomes of the deliberative ability is beyond the scope of this article. In the following, we only provide an exploratory analysis by linking individual deliberative ability to opinion change. Regarding opinion change, we focus on two preference items in connection with the immigration topic: "regularizing irregular immigrants" and the "degree of European involvement in immigration affairs". These two items formed the core of the Europolis immigration discussions and were, more or less intensely, discussed in all of our groups, which is essential when exploring the consequences of different deliberative abilities.[18] We evaluate both items at two time points, namely immediately before the deliberation phase (T2) and at the end of the Europolis event (T3). Regarding deliberative ability, we follow research on intelligence and IQ scores (see, e.g., Plomin et al., 1994) and take a special focus on low skilled vs. high skilled deliberators. Hereby, we consider the top and low ends of our latent variable of deliberative quality. Concretely, we selected participants above the 90^{th} percentile and below the 10^{th} percentile of our latent construct of deliberative quality (leading to 18 participants in each of the two groups). In so doing, we consider those participants whose ideal points were clearly different from each other in the item response analysis (see Figure 2.2). A glance at the raw data shows that participants in the group of low skilled deliberators reach a maximum of two standards of deliberative quality, whereas participants in the group of high skilled deliberators reach five or six quality standards.

Table 2.2 presents the raw data and the t-tests between the high and low skilled deliberators (and for comparison purposes, the raw data of the whole sample). Looking at the absolute change, we find that both high and low skilled deliberators change minds: the average change rates on the 11-point scales are 1.06 ("regularizing irregular immigration") and 2.50 ("European involvement") for high skilled deliberators and 1.94 and 2.76 for low skilled deliberators, respectively. Thus, at a very general level, deliberative ability is not a mere "etiquette". Both groups changed their minds and although high skilled deliberators did so to a (slightly) smaller degree, the difference is in neither case significant at a level of $p<0.1$. In sum, trends underline that we do not confront a "worst case scenario" where only the opinions of deliberatively low skilled and otherwise disadvantaged deliberators shift considerably

[18] The item "regularizing irregular immigrants" was measured on the basis of an 11-point scale where participants had to rank their policy choice between wholehearted agreement with return (0) and wholehearted agreement with legalization (10). The item "European involvement" was measured on the following question: On a scale from 0 to 10, where '0' means "entirely at the EU level", and '10' means "entirely by the individual Member States", and '5' is exactly in the middle", at what level do you think decisions should be made in each of the following areas? We inverted the scale so that higher values indicate more "European involvement".

while the high skilled sit on their positions and do not change minds. Interestingly, however, the two groups differ considerably in respect of *relative* attitude changes. While for the item "regularizing irregular immigration", high skilled deliberators display more "progressive" attitudes towards irregular immigrants after deliberation, low skilled deliberators have moved towards the "conservative" end of the scale. The complete opposite occurred for the "European involvement" item. While the two groups differ significantly and substantially in their pre-deliberative opinions, low skilled deliberators closed the gap by turning more European.[19] At this point, we do not have an explanation for the detected patterns. We regard the examination of deliberation's effect on post-deliberative opinions as an important issue for future research. Chapter 4 will make an initial attempt to explore this subject.

Table 2.2: Pre- and Post-Deliberative Opinions and Absolute Opinion Change

	Irregular immigration [b]			**European involvement** [c]		
	T2	*T3*	*Absolute change*	*T2*	*T3*	*Absolute change*
Whole sample (N=185)	5.57 (0.18)	5.72 (0.18)	1.58 (0.15)	4.84 (0.22)	5.06 (0.20)	2.33 (0.18)
High skilled (N=18)	6.11 (0.56)	6.83* (0.53)	1.06 (0.25)	6.22 (0.81)	6.06 (0.43)	2.5 (0.59)
Low skilled (N=17) [a]	5.94 (0.55)	4.82^{+} (0.61)	1.94 (0.54)	3.71 (0.67)	5.06 (0.66)	2.76 (0.63)
	Mean difference			Mean difference		
Between groups	0.17	2.01*	- 0.89	2.52*	1.00	-0.26

Note: Standard errors are shown in parentheses. Levels of significance are shown for within-group differences (first three lines) and differences between the high and low skilled group (bottom line). $^{+}$p<0.1, *p<0.05, **p<0.01. [a] Due to missing information in questionnaire responses, the low skilled sample is reduced by one participant for both items. [b] The question is captured on an 11-point scale, where 0 indicates full agreement with sending irregular immigrants back and 10 indicates full agreement with regularizing all irregular immigrants currently living in the country. [c] The question is captured on an 11-point scale, where 0 indicates that decisions on immigration should be made entirely at the national level and 10 indicates that decisions should be made entirely at the EU level.

Our findings have important implications for the prospects of global citizen democracy as well as the design of deliberative citizen events. Transnational citizen deliberation may be more fragmented than many deliberative democrats are willing to admit. This particularly concerns the various proposals for deliberative citizen involvements in global politics (Dryzek et al., 2011; Goodin and Ratner, 2011). Dryzek et al. (2011: 4), for instance, draw from non-

[19] However, the standard errors are large and therefore, the difference between high and low skilled deliberators is not significant.

Western experiences with citizen deliberation and assert that deliberation is not "the preserve of wealthy societies and educated participants". Hence, global citizen deliberation is presented as a manageable event that will have the capacity of bringing highly diverse citizens together rather than reinforcing already existing differences. But as the Europolis event makes clear, overly optimistic expectations that deliberation is an event where all participants engage equally and deliberate together at high quality levels are romantic dreams indeed. Even though deliberative ability is far from being socially or culturally determined, class and region turned out to be important predictors of deliberative performance. Moreover, biases are amplified by the fact that deliberatively high skilled participants not only speak more and longer, but also have different opinions at the end of the event. Therefore, proposals for deliberative citizen involvement in global or transnational affairs must make special provisions to properly integrate people from diverse social and cultural backgrounds, especially if these combine widely varying deliberative skills (see also Rosenberg, 2005: 222). In this regard, a more dedicated focus on alternative forms of communication – such as story-telling – may only be a partial solution for better integrating low skilled deliberators, since story-telling is a technique employed by high skilled deliberators as well. Another pathway might be to focus on group composition (and voting rules). Karpowitz and Mendelberg (2007) have demonstrated empirically that a high share of women and cultural minorities in discussion groups in combination with unanimity voting rules came closest to deliberative ideals on a range of possible outcome measures. Here, group composition (in combination with voting rules) altered the functioning of the whole group, empowering disadvantaged groups and equalizing deliberative interactions (see also Karpowitz et al., 2012). Thus, by creating a strong presence of disadvantaged people in discussion groups and by giving these groups a common decision task (e.g., working out a common statement; see Grönlund et al., 2010), inequalities in the process may be reduced. Finally, one might also conceive of a "learning phase" prior to the "real deliberation" phase (Kanra, 2005), where disadvantaged participants are given the possibility to familiarize themselves with deliberative repertoires (see also Mansbridge, 1996b: 57–58). In particular, we would like them to enhance their standing in deliberation, i.e., speaking up more frequently, as well as developing and cultivating their rhetorical skills. Only when this is the case can we expect that more egalitarian patterns of deliberative interaction evolve with advantaged participants seriously listening to the viewpoints and opinions of less advantaged participants.

2.7 Conclusion

This is one of the first studies exploring deliberative abilities of ordinary citizens on the basis of a philosophically grounded measure, an updated version of the Discourse Quality Index (DQI). The object of our study was Europolis, a transnational deliberative poll. The Europolis event yields a fairly nuanced and mixed picture of citizen deliberation. On the positive side, citizens' deliberative abilities are more widespread than assumed by skeptics and demanding standards of classic deliberation are far from being utopian standards that no citizen can achieve. Moreover, high skilled deliberators as envisaged by deliberative theory exist in the realm of citizen deliberation. An item response analysis shows that classic deliberative quality forms a latent dimension with participants scoring high on all deliberative standards, ranging from justification rationality to common good orientation, respect, empathy, and inquisitiveness. Finally, high deliberative ability is not fully determined by social or cultural characteristics: important social variables - especially gender - are not associated with deliberative ability. On the negative side, we nonetheless detected an important class-region bias in the ability of deliberating: compared to other participants, working-class participants from Central Eastern and Southern Europe were considerably less likely to reach the various standards of high quality deliberation. Moreover, putting our findings in perspective, we detect a slightly worrisome trends: our measure of deliberative ability is moderately correlated with the frequency and length of speaking, giving a preliminary indication that less deliberatively skilled participants find it more difficult to express their concerns. In sum, while citizen deliberation may work better than its most fervent critics have claimed, it is a much more fragmented exercise than its staunchest advocates are ready to admit.

Of course, our study is not without limitations. First, critics might say that our results and the implications we draw from them are driven by our conceptualization of deliberative quality. Had we set less demanding deliberative standards, then the biases would not have shown up and we could draw a much brighter picture of citizen deliberation. We disagree. On the one hand, deliberation – even in its re-configurations (see Bächtiger et al., 2010a) – is a demanding concept of political communication, and critics have attacked its viability not on watered-down variant but on the classic and Habermasian-inspired version thereof. On the other hand, even if one applies a more lenient version of deliberative quality, namely the Fishkinean conception of "reasoned and civil discussion", similar biases than we found for our classic and Habermasian-inspired conception of deliberation show up (analyses upon request). Second, one might object that we have only analyzed 13 out of 25 discussion groups

in Europolis. However, since there are no signs of systematic variation in maximum deliberative performance across discussion groups, we wonder how much empirical leverage will be gained by analyzing more groups. Third, some of the measures used in the analysis are not optimal and stronger associations might have been found if we had other measures. This mainly concerns psychological factors. There is new research linking personality traits to electoral behavior, political participation, and political attitudes (see Gastil et al., 2008; Mondak et al., 2010). It is urgent to include such variables in future questionnaires of citizen deliberative events as well. Fourth, our analysis only considers the *formal* aspects of deliberative quality while neglecting other aspects, namely the informational quality of the discussions, the substantive content of arguments as well as argumentative balance. Fishkin (2009: 34), for instance, emphasizes that argumentative balance, i.e., "the extent to which arguments offered by one side or from one perspective are answered by considerations offered by those who hold other perspectives", is a crucial element of deliberative quality. Future research will need to consider such measures of deliberative quality as well. And fourth, while we have made a first attempt at linking deliberative ability to opinion change, much more could and should be done. Indeed, the link between process quality and outcomes has been largely neglected in previous studies (for exceptions, see Neblo, 2007a; Baccaro et al., 2011; Sanders, 2012). Yet, optimistic conclusions for deliberative theory do not follow unless one can demonstrate that deliberation changes opinions "primarily via mechanisms specified in the normative theories" (Neblo, 2007a: 1). This means, on the one hand, that a high deliberative ability should be linked to the willingness of participants to change their opinions. On the other hand, we need to explore whether deliberative quality can serve as a resource, in that the provision of high quality arguments influence the direction of opinion shift – at least in a safe venue for deliberation such as Europolis. For an initial attempt to remedy this deficiency, see Chapter 4.

Despite these limitations, our study is the first one to perform a comprehensive, systematic, and in-depth analysis of deliberative abilities of ordinary citizens. At the same time, it provides researchers with a tool set that can be applied to an in-depth evaluation of the booming numbers of citizen deliberative events world-wide.

Chapter 3

EQUAL PARTNERS IN DIALOGUE? PARTICIPATION EQUALITY IN EUROPOLIS*

* I would like to thank André Bächtiger, Marc Bühlmann, Sandra Lavenex, Michael Morell, Michael Neblo, Seraina Pedrini, Marco Steenbergen, and Richard Traunmüller for valuable comments and advice on earlier versions of this paper. I am also grateful to participants at the panel "Transnational Citizen Participation in Europe" at the ECPR General Conference in Reykjavik 2011, to the participants at the panel "Discussion, Deliberation, and Public Opinion" at the MPSA Conference 2012, and to participants at the 2012 *Deliberative Democracy in Action* Conference in Abo, Finland, for useful remarks and thoughts.

A revised version of this paper has been published 2015 in *Political Studies 63(S1): 110–130.*

> "Deliberation depends on achieving an important kind of equality, equality as a partner in dialogue." *Warren (2001: 81)*

3.1 Introduction

By supporting diverse participatory approaches, Brussels has recently made efforts to bring the European Union closer to its citizens. Among these efforts, the pan-European deliberative poll (DP) on third country migration and climate change (*Europolis*) may be considered one of the most extensive attempts to increase citizens' involvement in European affairs. By gathering a representative sample of the European population for deliberation, Europolis created a favorable environment for EU-citizens of all 27 member states to become substantially more informed about politics (Fishkin, 2009: chp. 6). At the core of molding more informed opinions lies the premise of political equality, which is according to Fishkin and Luskin (2005: 285) the requirement for "equal consideration of everyone's preferences". Up to now, evaluations of political equality in DP have largely focused on the questions of whether the sample of deliberators was representative of a broader public (e.g., Andersen & Hansen, 2007; Fishkin et al., 2010; Luskin et al., 2002) and whether post-deliberative opinions were dominated by the view of a particular social group (e.g., Fishkin et al., 2010; Luskin et al., 2013).

Less is known about the degree to which participants actually gained equal standing in the process of deliberation itself (see Kim et al., 2010; Siu, 2009 for partial exceptions). This is a substantial shortcoming for at least two reasons. First, while offering everyone equal possibilities to participate in mini-public events is essential, it is in itself not a sufficient condition to establish equal participation (e.g., Bohman, 1997; Cohen, 1989; Fraser, 1992; Young, 2002). Critics argue that informal hurdles to participation can persist even after individuals gain formal access to the floor. However, organizers of mini-publics such as the DP mitigate this concern by emphasizing that mini-public designs incorporate many procedural safeguards: the introduction of trained facilitators and specific rules for discussion should ensure that deliberation proceeds in a fair and equal way (e.g., Fishkin and Luskin, 2005). Yet, it is surprising how little the role of the facilitators has been theorized in the framework of deliberation, given the essential part that is attributed to facilitators in establishing an inclusive atmosphere for discussion (Smith, 2009: chp. 6; see Moore, 2012 for an exception). It comes as no surprise that the lack of theorization echoes in largely mixed results on facilitators' behavior and their potential influence on establishing a balanced atmosphere for discussion (e.g., Ryfe, 2006; Schneiderhan and Khan, 2008; Stromer-Galley,

2007; but see Trénel, 2010). Thus, an evaluation of deliberative participation that only focuses on formal access to deliberation may be insufficient. Second, estimating whether post-deliberative opinions are strongly influenced by a particular social group may not give a full account of whether participants were substantively equal in deliberation. Of course, confirming that disadvantaged groups were not forced to conform to the views of a dominant group by underlying power structures is a positive sign of deliberative inclusion. Yet, an outcome-oriented approach to participation equality might be too quick in assuming that serious reflection of others' arguments automatically manifests in opinion shift (Dobson, 2012: 844; see Barabas, 2004). In other words, whether others' thoughts manage to change one's opinion is, ideally, more a question of the force of the better argument than of the actual willingness to seriously consider others' claims in the light of one's own (Mansbridge, 1999: 225). Moreover, DP is an *educative forum* and its prime aim is the formation of considered opinions (Fung, 2003). In order for citizens to get enlightened preferences, a deliberative process has to ensure that the relevant reasons are aired and appropriately weighed (Mansbridge et al., 2012). Only a focus on the process of deliberation allows for testing whether this was actually the case.

This article adds another layer to the literature on inclusiveness of deliberative mini-publics by analyzing participation equality in the process of deliberation itself. Specifically, the article argues in favor of two additional dimensions of substantive participation equality that are regarded as crucial preconditions to creating more informed and considered opinions. I dub them *equal contribution* and *equal consideration*. From such a perspective, I address the questions of the extent to which individuals participated at equal levels and whether their arguments were considered on equal merits. I propose multilevel modeling and social network analysis as possible methodological tools to address these questions. A subsample of Europolis small group discussions on third country migration serves as a case study.

The remainder of the article is organized as follows. The next section contains the theoretical discussion on equal opportunities to influence the deliberative process and embeds the argument in deliberative theory. Section 3.3 displays the empirical model and details the statistical methodologies at use. Subsequently, I present the main results (section 3.4) and discuss the implications of the findings (section 3.5). Section 3.6 concludes and offers suggestions for future research.

3.2 Participation Equality in Deliberation

Democratic equality has many facets (e.g., Dahl, 1989, Saward, 2003) and theorists also disagree about the kind of equality that deliberation should be concerned with (see Thompson, 2008). A broad and often cited definition stems from Joshua Cohen (1989). Cohen (1989: 22-23) states that not only do individuals need to be formally empowered to participate in deliberation, they also need to be substantively equal "in that the existing distribution of power and resources does not shape their chances to contribute to deliberation, nor does that contribution play an authoritative role in their deliberation". Although random sampling methods as utilized in DP may succeed in gathering a representative sample of a broader public and therefore fulfill a necessary condition for equal participation, they are in themselves not sufficient.[20] Equalizing the chances to take part in DP is what Young (2002) refers to as external inclusion. The concern about remaining informal hurdles to effective participation, however, is related to her conception of internal inclusion. According to Young (2002: 55), internal inclusion is violated when some people or particular groups lack effective opportunities to influence the political decision-making process, despite having gained access to the political floor. Young (2002) fears that traditional power structures will be transferred to deliberative settings because the calm, reasoned and dispassionate style of expression that classic deliberation demands only suits the already advantaged (see also Sanders, 1997).

According to Knight and Johnson, 1997: 292), lacking equal opportunities to political influence is a violation of "the fundamental notion that democratic outcomes are the product of the interest of equal citizens". By viewing deliberation as "uncoerced give and take of reasoned arguments" (p. 281), the authors are among the few to explicitly take into account the dialogical element of deliberative participation. In such a view, a speaker's opportunity to political influence is determined not only by the action of the speaker himself but also by the listeners' willingness to take the speaker's claim into account (see also Bohman, 1997). Yet the authors advance a capacity-based conception of equal opportunities that, first and

[20] There are a few reservations about the benefits of statistical representativeness. Comparing the sample of participants to the control sample neglects another level of possible exclusion that may have occurred before the sample was gathered. On the one hand, such exclusion might be technical: random digital dialing bars people without landline from participation, and these are predominantly young people (see Hansen, 2004; chp. 2). On the other hand, exclusion might also be systematic in that those unwilling to take part in the initial survey are likely to differ in some important ways from the ones that demonstrate high motivation for participation (Parkinson, 2006a: chp. 4). Moreover, statistical representation is not the same as sampling an adequate cross-section of the population (Parkinson, 2003; Brown, 2006), nor does it guarantee discursive representation (Dryzek and Niemeyer, 2008).

foremost, requires everyone to possess equal abilities to reason and articulate ideas in order to influence political outcomes (Knight and Johnson, 1997: 295-299).

However, deliberative theory would meet such an approach to equal opportunities with reticence. *A priori* denying a particular social perspective or non-rational styles of communication from exerting political influence would be counterproductive to the goal of deliberative inclusion (e.g., Bohman, 2006; Cohen, 1989: 30). By enlarging the pool of admissible forms of deliberative communication such as rhetoric or storytelling, deliberative theory recently leveled the requirement for rational discourse, which gave pause to theorists calling for equal reasoning capacities (e.g., Bächtiger et al., 2010a; Dryzek, 2000; see also Mansbridge, 1999: 233). To be sure, in order to influence deliberative outcomes, political equality requires that deliberators possess at least basic deliberative abilities for effective participation. Most deliberative theorists would also agree that deliberative outcomes should be guided by epistemic principles rather than requiring everyone to exert equal influence on the outcome (e.g., Estlund, 2000, 2009; Knigth and Johnson, 1997; Mansbridge, 1999). As Estlund (2000: 138) holds, whether one's arguments have been treated with adequate respect is not reflected in equal influence on the outcome, nor are inequalities in the outcome a clear indicator of disrespect. In this sense, Bohman (1997) presents a more convincing approach to equal opportunities that emphasizes the requirement for effective social freedom, i.e. the willingness of others to recognize one's claim as a possible relevant reason. Bohman (1997: 345) adheres that "citizens will then know when their failure to achieve their political goals has to do with their lack of convincing reasons rather than limits upon the extent of their effective social freedom or the greater freedom of others". In his argumentation, Bohman makes an important distinction between being able to influence the *process* of deliberation and the ability to influence the *outcome* of deliberation (or, the ability to *effectively* influence the deliberative process). While the former requires active communication of one's central claims and others' willingness to weigh the merits of each claim on equal grounds, the latter additionally requires one's ability to convince others by superiority of reasoning. The scope of this article is limited to the former.[21] The following two subsections elaborate on the theoretical foundations and practical limitations of such an approach.

3.2.1 Equal Contribution

From an epistemic point of view, deliberation is desirable because it holds the potential for incorporating a wide array of different perspectives (e.g., Bohman, 2006; Estlund, 1997;

[21] But see Chapter 4 for an initial attempt to capture the latter.

Mansbridge et al., 2012). By gathering a random sample of a given population, DP possesses the potential to expose participants to a wide range of facts and moral positions which may render individuals more considerate, further tolerance and may even reduce prejudice (Mutz, 2006). However, in order to unleash its full potential, deliberation requires everyone to express his preferences in a free and unconstrained manner (Cohen, 1989: 29). Sure, the deliberative ideal of equal participation does not demand that everyone participates with equal numbers of words (Mansbridge, 1999: 232). However, equal participation requires that no person or particular group completely dominates the discussion, even though deliberators may differ in power and prestige (Thompson, 2008: 501). Patterns of dominance in discussion would have consequential framing and agenda-setting effects and preclude people from basing their judgments on the entire pool of relevant reasons (see e.g., Mendelberg, 2002; Schulz-Hardt et al., 2000).

Several scholars express reservations when it comes to the practicability of equal contribution (e.g., Benhabib, 1996; Sanders, 1997; Williams, 2000; Young, 2000). This is based on the fact that these critics attribute a calm and dispassionate style of expression to middle- and upper-class white men, while they view the emotive and physically demonstrative modes of expression to be strongly associated with women and cultural minorities. As a consequence, Young (2002) fears that historically oppressed segments of the population do not dare to speak, as they do not consider themselves able to fulfill the requirement of rational discourse. As a possible cure, Mansbridge (1996b: 57-59) suggests that enclave deliberation among the disempowered should precede more heterogeneous deliberation. While deliberating among themselves, the less powerful could elaborate on their ideas and positions in a secure environment and prepare to confront the dominant groups with sufficient confidence (see also Karpowitz et al., 2009). Yet some researchers would contest the concerns related to gender by arguing that women are particularly prone to deliberation, in that they pursue a kinder and gentler style of politics (Childs, 2004; Norris, 1996; see also Karpowitz and Mendelberg, 2007). Moreover, Neblo et al. (2010) provide convincing evidence that groups that normally display reluctance towards traditional means of political participation are actually more willing to participate in deliberation. Hoping to get a better hearing of their claims, these groups might view deliberation as a valuable alternative to “politics as usual”. Whether mini-publics manage to activate disadvantaged groups to a substantial degree is still an open question and the few studies that addressed this subject so far provide mixed evidence (Hansen, 2010; Himmelroos, 2011; Kim et al., 2010; Siu, 2009).

3.2.2 Equal Consideration

By filtering incoming information in a biased way, deliberators may infringe another form of equality. For deliberation to reach its full epistemic potential, post-deliberative preferences and decisions need to be based on facts and logic and be the outcome of "substantive and meaningful consideration of relevant reasons" (Mansbridge et al., 2012: 11). In order for relevant reasons to enter the opinion formation process, deliberators' claims need to be listened to seriously (e.g., Bickford, 1996; Dobson, 2010; Mansbridge et al., 2012). However, deliberative researchers often ignore the importance of listening in that they often (and against all evidence) assume that incoming information is processed in a uniform way (Dobson, 2012; Lupia, 2002; Mendelberg, 2002). From a viewpoint of equal opportunities to influence the deliberative process, however, it is essential that everyone be recognized as equal discussant with legitimate claims (e.g., Bohman, 1997; Knight and Johnson, 1997). Moreover, a demanding understanding of equal consideration requires recognition to go beyond the mere and rather passive act of listening to what others have to say. As Bickford (1996: 153-159) succinctly puts it, silence may either be an indicator of attentive listening *or* a non-verbal refusal to engage with what others have said. In other words, deliberation requires uptake of and engagement with others' claims, so that people respond to claims "as issues – that is, as matters that deserve response in their own terms" (Warren 2001: 81; see also Goodin, 2000: 90-92; Pedrini et al., 2013). The very interactive moment of reciprocal recognition is also embodied in Gutmann and Thompson's (1996: 79) definition of mutual respect, which demands a "favorable attitude toward, and constructive interaction with, the persons with whom one disagrees". According to the authors, a listener's willingness to treat a speaker with respect also demonstrates his commitment for self-reflection and openness to preference change. And it is only to the extent that the listener *actively responds* to what a speaker has said that everyone, including the speaker, is able to reflect on public deliberation (see Christiano, 2008: 201).

Again, critics question the direct applicability of such a deliberative ideal to real-world deliberation. One argument is that dominant groups may, even unconsciously, treat claims of people from traditionally disadvantaged groups with inadequate respect (Sanders, 1997; Young, 2002). If true, this would dispel many of the benefits accredited to deliberation: "when this happens, democratic theory doesn't have an answer, because one cannot counter a pernicious group dynamic with a good reason" (Sanders, 1997: 354). Adequate consideration of others' views may also be more difficult to achieve when the group increases in heterogeneity. By bringing together people with different cultural backgrounds, more time

and energy might be required in order to appropriately consider the full variety of speaking and reasoning styles (Benhabib, 2002: 138-139; Burkhalter, 2002). The well-known insights provided by the interracial town meetings on school desegregation amplify some of these concerns (Mendelberg and Oleske, 2000). On the other hand, more optimistic conclusions are drawn when it comes to the uptake of claims from language minorities' in the Swiss parliament (Pedrini et al., 2013). Beyond these valuable insights, we lack empirical observations on the practicability of equal consideration. In the transnational environment of Europolis, the viability of equal consideration may particularly be endangered by the fact that people enter the discussion with a variety of cultural backgrounds and experiences in the field of EU-politics and the topic of immigration (e.g., Freeman, 1995; Geddes, 2003). When it comes to deliberative polling, Fishkin (2009: 160) admits that the degree to which equal consideration is realized cannot be prescribed by institutional design. To date, we lack empirical investigations on equal consideration in DP.

3.3 Data and Measurement

3.3.1 Case Study (Europolis)

Empirically, I analyze the discussions on third country migration in a subsample of 13 Europolis discussion groups (see section 2.4.1). Due to limitations in simultaneous translation, national heterogeneity of Europolis small group compositions varied to a considerable degree: while some groups encompassed a wide variety of nationalities, in others it merely came down to a discussion among neighbors. The Europolis project thus presents an interesting case for analyzing small group deliberation with varying degrees of European diversity. Although the subsample is limited in scope with regard to EU member states, it is comparable to the rest of the sample with regard to important socio-demographic and psychological variables (see Appendix D). Despite its limitations concerning the representation of different nationalities, the sample consists of groups that show notable variation in national heterogeneity.

3.3.2 Statistical Analysis

As individual participants are embedded in small groups for discussion, multilevel modeling seems the appropriate approach for the case of Europolis (Mendelberg, 2005: 641-642). The central idea behind a multilevel approach is that institutional or other context-level determinants do have an influence on individual level phenomena, but their influence is not necessarily determinant or uniform (see, e.g., Steenbergen and Jones, 2002). By estimating

random effects, multilevel modeling acknowledges that the estimated predictors may not account for all variation that exists between the groups. Given that exploring social group dynamics is still a lacuna in deliberative mini-public research, multilevel models may offer a welcome opportunity to account for this situation in a proper way.

The previous reluctance to use multilevel modeling in small group discussions might stem from the fact that the number of deliberating groups is usually rather low. Yet Gelman and Hill (2007: 247) assert that a model with as little as five groups may already be suitable for multilevel analysis. Nevertheless, there is the danger that using maximum likelihood inference, variance components are not properly estimated and estimates and confidence intervals can be biased, particularly when cross-level interactions are estimated in models where the number of level-2 units is smaller than 15 (Stegmueller, 2013; see also Browne and Draper, 2006). Due to the small number of groups in our sample (N=13), only rather simple models will be estimated. By assigning each individual to a particular group for discussion, the hierarchical structure in DP is straightforward, allowing for less complex models to be run. In order to estimate more complex models, fully Bayesian procedures would be appropriate (see, e.g., Hangartner et al., 2007; Bächtiger and Hangartner, 2010).[22] Where cross-level interactions are estimated, an informal approach to Bayesian inference complements the analyses: model simulations are run in order to retrieve an uncertainty distribution of the parameter estimates. The multilevel models run in this article are estimated using maximum likelihood estimation (MLE) with the lme4 package in R (Bates et al., 2011).

Since deliberation is a communicative act and an "individual functions as a part of the group and each influences the other" (Mendelberg, 2005: 645), social network analysis may help to capture elements of deliberative inclusion. Yet, with a few exceptions, network structures of deliberative discussions remain largely unexplored (see Cinalli and O'Flynn, 2013; Graham and Witschge, 2003 for exceptions). In order to explore consideration structures, network measures of homophily (e.g., McPherson et al., 2001) and centrality (Borgatti, 2005; Freeman, 1979) may be of particular relevance. Centrality measures capture an actor's connectedness to a set of other actors and are commonly viewed as an indicator of an actor's power within the network (see, e.g., Krackhardt, 1990; Ibarra, 1993). In deliberative

[22] Contrary to Hangartner et al. (2007) who analyzed deliberative behavior in the parliamentary setting, I have to deal neither with cross-classification nor with multiple hierarchical levels, which would increase model complexity to a considerable degree. Yet, due to the low number of groups, only models with one varying component (random intercept or random slope) are estimated. Particularly in models with cross-level interactions, random intercept models represent a simplification in assuming that individual differences in the outcome variable are completely accounted for by the moderating variable.

discussions, an actor's power may be determined by the degree to which his views have been taken up by his co-discussants. In order to evaluate whether participants' claims have received equal consideration, the dependent variable will be operationalized as a measure of *consideration centrality* (see below). Yet an empirical approach to equal consideration not only requires that claims are taken up, but also that the claims of some groups have been considered by other groups (Pedrini et al., 2013). Finding that members of some groups only talk to each other may have the same pernicious consequences as enclave deliberation without follow-up deliberation in more heterogeneous groups. As Mansbridge (1996b: 58) puts it: "when white supremacists only speak to white supremacists, Serbs to Serbs, feminists to feminists and political philosophers to political philosophers, they encourage one another not to hear anyone else". In this sense, homophily analysis may serve as an important complement to the analysis of consideration centrality. Homophily is defined as the tendency for similar individuals to form a relation (McPherson et al., 2001). Translated to the deliberative setting, homophily allows for testing whether members of a dominant group remained sufficiently open to claims raised by more peripheral actors or whether they preferred to engage with dominant others and thus contributed to the monopolization of the discussion.

3.3.3 Dependent Variables

In order to capture the essentials of the two dimensions of procedural equality, I rely on two features of the updated version of the Discourse Quality Index (DQI) (Bächtiger et al., 2010b). The DQI is a content-analytic measure that attempts to capture the essentials of Habermas' discourse ethics at the level of the speech act (Steenbergen et al., 2003). Originally, the DQI was developed to investigate parliamentary debates (Bächtiger, 2005; Steiner et al., 2004). With the recent amendments, however, the DQI has been applied successfully to the evaluation of citizen deliberation.

Contribution: Since there is no universal agreement on how to capture contribution, a double-track approach is pursued (see also Stromer-Galley, 2007). A participant's contribution is captured via the amount of speaking time (in seconds) and the frequency of contribution.[23] Linear multilevel models are estimated for the logarithm of speaking time. The distribution of speaking time is strongly skewed to the left and ranges from 0 to 1593 seconds (26.5 minutes) of speaking time in a total of approximately 1.5 hours of discussion time. Random slopes

[23] The original DQI captures equality by looking at whether participants were hindered from speaking freely by interruptions from others. In Europolis, interruptions were technically impossible due to simultaneous translation. Therefore, contribution in Europolis has only been assessed via the updated measure of the DQI.

instead of random intercept models are estimated since overall speaking time was roughly equal in all groups. The slopes for participants of EU-members that joined the EU in 2004 or later were allowed to vary. The underlying assumption is that members of new EU-countries show a more balanced participation behavior if they discuss only among themselves. The distribution of frequency is also strongly skewed to the left, ranging from 0 to a total of 21 delivered speeches. Here, Poisson models for count data were estimated. In order to account for overdispersion, Poisson lognormal models are estimated by adding an individual-level random effect to the models estimated with lme4 (Bates et al., 2011; Bolker, 2010).

Consideration is based on the updated DQI indicator of respect towards other participants' arguments. The DQI indicator captures interactivity, or the uptake of arguments, in assessing whether speakers explicitly referred to others' arguments in a disrespectful, neutral or respectful manner or whether they ignored other participants' statements. An explicit reference is present if someone directly addresses a previous speaker in response to something he had said. In a first step, interactive instances have been coded from the transcripts of the group discussions. An inter-coder reliability test by three independent coders showed respectable levels of agreement on the coding.[24] By assessing the nature of the reference as well as the sender and the receiver of a reference, a matrix of argumentative uptake can be created (see Pedrini et al., 2013). This presents the basis for further analysis. In this paper, I focus on explicitly neutral or positive references and subsume them under the category of "civil references". I ignore disrespectful references, since they violate the fundamental notion of constructive interaction (Gutmann and Thompson, 1996; Mansbridge et al., 2012).[25] Moreover, instances of disrespect were rare in Europolis. This is not surprising given that in DP, stakes are low and participants are encouraged to meet others with respect (Fung, 2003). In order to retrieve the measure for consideration, an actor's centrality has been calculated from the matrix. Here, I focus on in-degree relations, i.e. the amount of references that a participant *receives*. Concretely, the network centrality measure encompasses the total number of co-discussants that made a civil reference to a speaker at least once. In the Europolis subsample, in-degree centrality ranges from 0 (a person spoke but has never been referenced) to 10 (a person was referenced by 10 different co-discussants at least once). Since the distribution is strongly skewed to the left, lognormal Poisson models are estimated (see above).

[24] See Appendix B.

[25] From a deliberative standpoint, however, receiving disrespectful feedback may be preferable to being completely ignored. I thank Michael Neblo for highlighting this point.

3.3.4 Predictor Variables

The following variables of interest were included in the model in order to test whether participation in Europolis was biased in relation to important demographic and socio-economic indicators: a binary variable for gender, taking in the value (1) for women and (0) for men; another binary variable indicating whether a person feels that she belongs to the working-class (1) or to the middle or upper class (0); a variable for education, measured in years of age when the education ended; and a categorical variable for age, classifying a person into being young (18-34 years old), intermediate (35-64 years old) or older (65 and above).

Two additional variables were introduced in order to account for the impact of national diversity. At the individual level, an additional categorical variable distinguishes between people belonging to a Western European (0), a Southern European (1) or a Central Eastern European state that joined the EU in 2004 or later (2). At the group level, a measure was created to capture the degree of national heterogeneity that was present within the groups. The measure is based on the *Ethno-Linguistic Fractionalization Index* (ELF) (see Fearon and Laitin, 2003). The ELF captures fractionalization based on the Herfindahl concentration index.[26] The original ELF can take in any value between zero and one and indicates the probability that two randomly selected individuals from the entire population will be from two different groups. Yet the ELF indicator does not take into account another crucial feature of European diversity, namely the broader cultural background of participants and the possibility that experiences with immigration differ across countries or regions (see Freeman, 1995; Geddes, 2003). Therefore, the group scores obtained on the basis of the ELF indicator have been weighed by a factor of three for groups that were composed of Western, Central Eastern, and Southern Europeans and by a factor of two for groups composed of Western and Central Eastern Europeans or groups composed of Western and Southern Europeans.[27] Groups where all participants originated from the same region obtained a weight of 1. The scores ranged from 0.48 to 2.11, with higher scores indicating higher levels of national heterogeneity.

Additional individual level factors are introduced as controls. All of them are expected to enhance an individual's engagement in discussion. Interest in politics is captured on an 11-point scale from (0) "not at all interested in politics" to (10) "passionately interested in politics". An indicator for salience aims to capture the personal relevance of the topic by

[26] $ELF = 1-\Sigma s_i^2$ where s_i is the share of group i out of a total of n groups.

[27] None of the groups of our subsample exclusively incorporated participants from Central Eastern and Southern Europe.

asking how immigration is perceived on a continuum from "no problem at all" (0) to "the most serious problem we face" (10). In order to model expertise in the field, a knowledge index was created that takes in a discrete value from 0 to 3, by indicating the number of correct answers to knowledge questions on immigration in the questionnaire.[28] Last but not least, a categorical variable for a person's attitude toward immigration is introduced. The variable has been built as a scale from five central questions on immigration that proved to be substantially related in a test of inter-item correlation.[29] The 25% at the lower end of the scale were classified as holding strong conservative opinions towards immigration, while the top 25% of the scale were classified as holding strong progressive opinions towards immigration. The in-between 50% represent the baseline category. All control variables express participants' answers at wave 2 of the questionnaire, i.e. right at the start of the small group discussions.[30]

3.4 Results

The first step was to estimate multilevel models for contribution. Model 1 in Table 3.1 presents the results of the lognormal Poisson models for frequency of participation. According to expectations, interest in politics is a clear driver of contribution in terms of utterances.

[28] The questions asked for the definition of Blue Card workers, the role of the EU in the current immigration policy and for a figure of the number of immigrants currently living in the EU. For each of the questions, the participants could choose out of four possible options and an additional "don't know" category.

[29] α=0.76. See Appendix D for more details on the construction of the scale.

[30] I refrain from inserting a group level predictor for facilitator intervention. First of all, DP is not designed to test the influence of different facilitation styles. Therefore, encountered differences in facilitator behavior are non-systematic and a categorization of different intervention styles proved difficult. Second, facilitator intervention *should be driven by participation patterns*. Only if participation is largely unequal, facilitators should interfere. If individuals participate on roughly equal levels from the start, there is no need for intervention. Therefore, a proper measure of facilitator behavior would require identifying cases where the necessity for facilitator intervention existed but where facilitators failed to comply with their tasks. Creating such a measure proved to be exceedingly difficult.

Table 3.1: Determinants of Contribution (incl. Silent Participants)

	Model 1 Frequency	*Model 2 Speaking Time*
Fixed effects		
Intercept	1.563*** (0.166)	5.621*** (0.245)
Women	-0.111 (0.116)	-0.663** (0.223)
Education	0.013 (0.013)	0.021 (0.026)
Working-class	-0.261^{+} (0.150)	-0.312 (0.280)
Younger participants	-0.144 (0.152)	0.018 (0.292)
Older participants	-0.131 (0.148)	0.150 (0.290)
Southern Europeans (SE)	-0.054 (0.182)	0.111 (0.320)
Central Eastern Europeans (CEE)	-0.357 $^{+}$ (0.188)	-0.783 (0.480)
Interest in politics	0.072* (0.030)	0.184*** (0.056)
Knowledge	0.038 (0.063)	0.301* (0.127)
Salience	-0.016 (0.027)	-0.003 (0.053)
Attitudes (strongly progressive)	0.073 (0.146)	-0.239 (0.289)
Attitudes (strongly conservative)	0.120 (0.148)	-0.447 (0.281)
National heterogeneity of small group	-0.306 (0.266)	0.001 (0.287)
Random effects		
Individual level	0.244 (0.494)	1.942 (1.394)
Group level	0.385 (0.148)	0.770 (0.877)
AIC	364.4	655.8
Log Likelihood	-166.2	-311.9

Notes: Multilevel models are estimated with the lme4 package in R (Bates et al., 2011). For contribution in speaking time, a linear hierarchical model with varying slopes for CEE participants is estimated. The dependent variable is operationalized as log of the total amount of speaking time. For frequency of contribution, a lognormal Poisson model is estimated (varying intercept) in order to account for overdispersion (Bolker, 2010; Bates et al., 2011). All ordinal and continuous predictor variables are centered at the mean. Estimated standard errors are shown in parentheses. $^{+}$p<0.1, *p<0.05, **p<0.01., ***p<.001. Individual N=173. Group N=13.

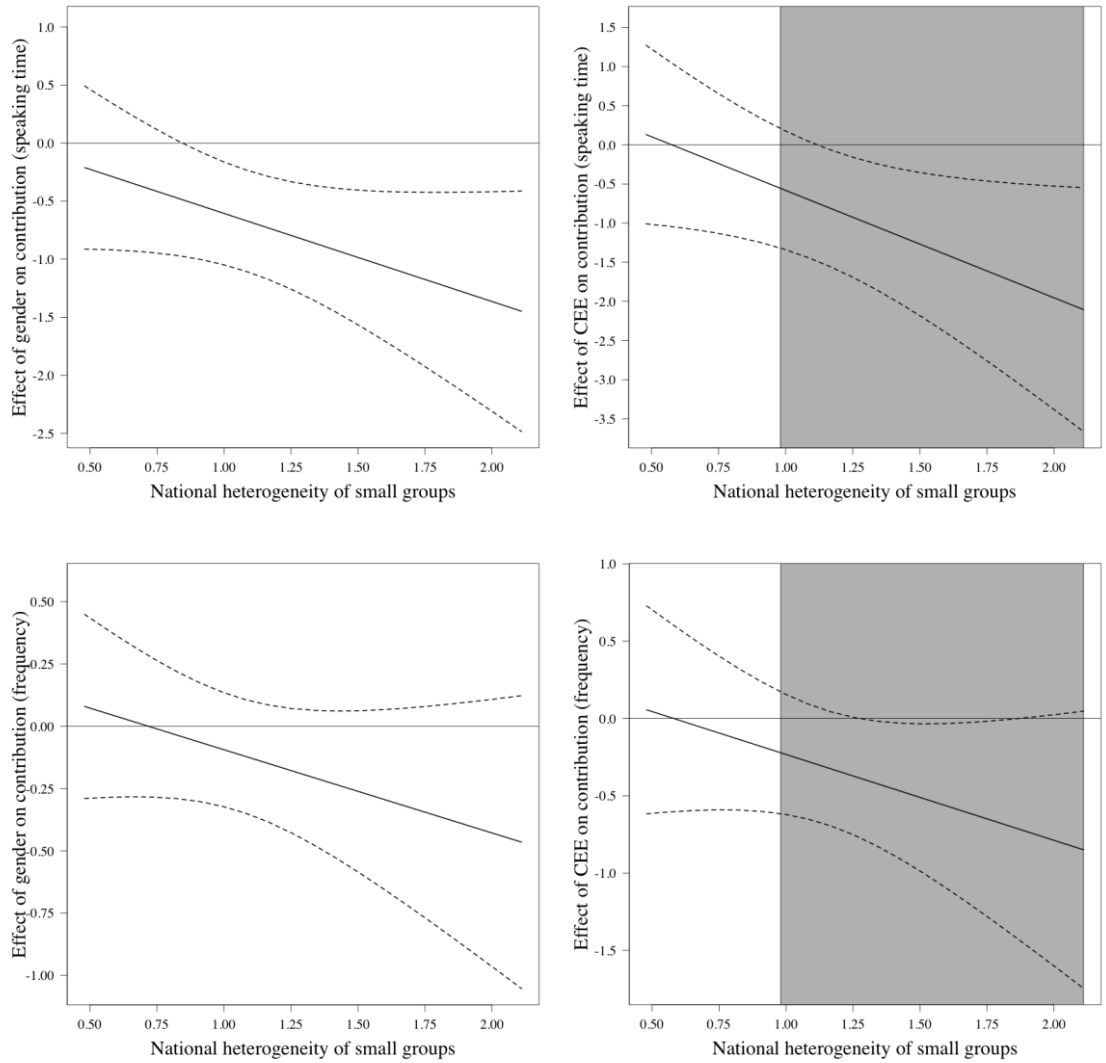
Exponentiating the estimates reveals that a one-point increase on the 11-point scale accounts for a significant difference in contribution of 7.5%, holding all other factors constant. In turn, belonging to a Central Eastern European (CEE) member state is supposed to reduce participation by 30% of the expected count. A strong reduction is also suggested for members of the working-class (-23%). However, there is still a considerable degree of statistical uncertainty in the latter two estimates ($p<0.1$). On the other hand, no bias is detected for age, education, or gender. Yet, the picture looks somewhat different when focusing on speaking time, the second indicator of contribution equality (Model 2, Table 3.1). Exponentiating the estimate for women shows that the significant reduction in speaking time is considerable (-48.5%), holding all other factors constant. Even higher is the estimated reduction in speaking time for participants from CEE states (-54.3%), although the effect slightly misses the level of marginal significance ($p<0.1$). While working-class participants are biased when it comes to frequency of speaking, there is no significant evidence that this bias translates to the amount of speaking time.

One fear of deliberative critics is that participation biases increase with the heterogeneity of the discussion environment (e.g., Karpowitz et al., 2009). This is explicit for women and participants from new EU-member countries.[31] Figure 3.1 displays the marginal effect plots based on the interaction models presented in Appendix E. Again, the bias is more pronounced for the models of speaking time while the effect for frequency of contribution does not reach sufficient levels of statistical significance. In terms of speaking time, women's contribution was significantly and substantially different from zero in the nine most heterogeneous discussion groups, while the gender gap is smaller and less certain in the more homogeneous groups. More worrisome is the picture for participants from new EU-member countries. The shaded area demarcates the location of the groups in which participants of the new CEE member states were engaging with participants from older EU-member states. From a perspective of European integration, the results are somewhat troubling: when engaging with participants from older EU-countries, the newcomers are very likely to add significantly less to the discussion. As for women, the gap in speaking time increases with national diversity. Since CEE members were more likely to be located in highly heterogeneous group discussions, it is reasonable to check for individual level interaction effects as well: although CEE women are slightly more likely to remain silent than Western European (WE) women,

[31] For working-class and Southern European participants, national heterogeneity did not prove to be an amplifying factor of inequalities (results available on request).

active CEE women are contributing more compared to active WE women, both in terms of frequency and speaking-time.[32]

Figure 3.1: The Effect of Women and CEE Participants on Contribution by National Heterogeneity of Small Groups



Note: Marginal effect and 95% confidence intervals are based on the hierarchical models presented in Table E.1 in Appendix E. The shaded areas demarcate the location of the groups in which CEE participants were engaging with participants from older EU-member states (WE, SE, or both). See also Figure E.1 (Appendix E) for the results of the simulation of the conditional effects.

That women participate considerably less compared to their male counterparts reflects findings from other face-to-face mini-public settings (Hansen, 2010; Himmelroos, 2011; see also Mansbridge, 1983) but contradicts findings from online deliberative polls (Kim et al., 2010; Siu, 2009). Yet in the Danish deliberative poll, Hansen (2010) concludes that the gender gap is particularly pronounced among the most passive speakers. The findings are

[32] The latter also translates to consideration. Results available on request.

similar in Europolis: Table 3.2 presents the contribution models by only including the active speakers, i.e. the participants that contributed at least once to the discussion. These results are important to keep in mind when turning to the consideration models, since active contribution is a necessary precondition for receiving consideration.

Being female results in a reduction in speaking time of 29%, which is considerably lower than the estimate of the models including the non-speakers, suggesting that a significant majority of the non-speakers are female. On the other hand, the effect for working-class participants increased both in magnitude and clarity. Among the active speakers, working-class participants seem to be the ones with the smallest amount of contribution. Holding all other factors constant, being a member of the working-class is associated with a substantial reduction of 48.7% in speaking time (and 28.6% in contribution frequency). With regard to frequency of contribution among the active speakers (Model 4), new biases appear for younger (-21.5%) and older participants (-20.7%) on a marginal level of significance ($p<0.1$). On the other hand, being a participant with strong conservative attitudes towards immigration increases the frequency of contribution by 28.8% ($p<0.1$). No such effect is detected for participants that display highly progressive viewpoints towards immigrants.

Let us now turn to the determinants of consideration (Table 3.2, Model 3). In terms of biases for our marginalized groups, the consideration model seems to be somewhat a reflection of the frequency model for active speakers presented in the same table (Model 2), indicating that the ones who spoke more often were also the ones whose views were considered more often by others.[33] With a value of 0.73, the bivariate correlation between consideration centrality and frequency of speaking is indeed considerable. According to the estimates, being a member of the working-class accounts for a 39% decrease in receiving consideration. For women and new EU-members, results look more promising: while there is no evident discrimination in terms of contribution frequency among the active speakers, there is also no evidence for their marginalization in terms of consideration.

[33] However, this is only true for the variables of interest. Turning to controls, it becomes evident that although interest in politics is a strong driver of participation, it is not an explanatory factor for receiving consideration. On the other hand, the ones that clearly receive more consideration compared to their rate of contribution are holders of conservative attitudes towards immigration.

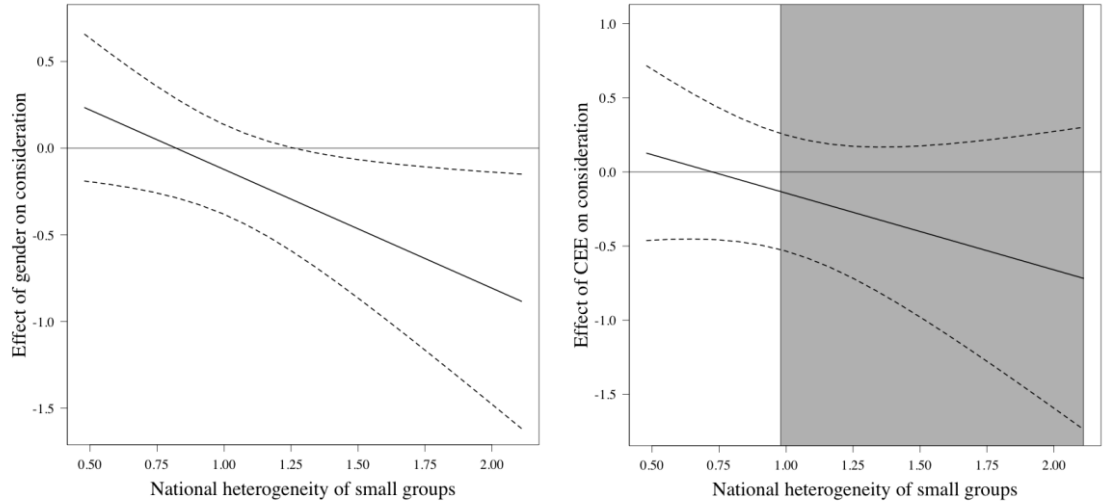
Table 3.2: Determinants of Contribution and Consideration (Active Speakers only)

	Contribution		**Consideration**
Fixed effects	*Model 3 Time*	*Model 4 Frequency*	*Model 5 Centrality*
Intercept	5.527***	1.589***	0.962***
	(0.415)	(0.160)	(0.152)
Women	-0.343*	0.001	-0.145
	(0.158)	(0.106)	(0.133)
Education	0.008	0.013	0.012
	(0.018)	(0.011)	(0.014)
Working-class	-0.667***	-0.337*	-0.494**
	(0.196)	(0.136)	(0.181)
Younger participants	-0.237	-0.242^{+}	-0.150
	(0.204)	(0.139)	(0.174)
Older participants	-0.238	-0.232^{+}	-0.298 $^{+}$
	(0.202)	(0.134)	(0.171)
Southern Europeans (SE)	0.119	-0.021	-0.307
	(0.220)	(0.166)	(0.203)
Central Eastern Europeans (CEE)	-0.209	-0.230	-0.185
	(0.270)	(0.174)	(0.201)
Interest in politics	0.108**	0.052^{+}	0.006
	(0.039)	(0.027)	(0.034)
Knowledge	0.110	-0.015	0.006
	(0.088)	(0.057)	(0.072)
Salience	-0.010	-0.014	-0.016
	(0.037)	(0.024)	(0.030)
Attitudes (strongly progressive)	-0.074	0.160	0.153
	(0.199)	(0.130)	(0.164)
Attitudes (strongly conservative)	-0.048	0.253^{+}	0.390*
	(0.200)	(0.135)	(0.163)
National heterogeneity	0.065	-0.279	-0.219
	(0.193)	(0.271)	(0.207)
Random effects			
Individual level	0.898	0.154	0.161
	(0.948)	(0.393)	(0.401)
Group level	0.133	0.166	0.047
	(0.365)	(0.407)	(0.217)
AIC	501.8	301.5	282.5
Log Likelihood	-234.9	-134.8	-125.2

Notes: Multilevel models are estimated with the lme4 package in R (Bates et al., 2011). For contribution in speaking time, a linear hierarchical model with varying slopes for CEE participants is estimated. The dependent variable is operationalized as log of the total amount of speaking time. For frequency of contribution and consideration, lognormal Poisson models are estimated (varying intercept) in order to account for overdispersion (Bolker, 2010; Bates et al., 2011). All ordinal and continuous predictor variables are centered at the mean. Estimated standard errors are shown in parentheses. $^{+}$p<0.1, *p<0.05, **p<0.01., ***p<.001. Individual N=163. Group N=13.

Again, interaction models were estimated. Figure 3.2 plots the marginal effects for gender and CEE members on receiving consideration. There is no concrete evidence that new EU-members received significantly less consideration than their older European peers. Yet, the confidence intervals for women in the most heterogeneous groups are different from 0, suggesting that women receive significantly less consideration in when they engage in highly heterogeneous groups.

Figure 3.2: The Effect of Women and CEE Participants on Consideration by National Heterogeneity of Small Groups

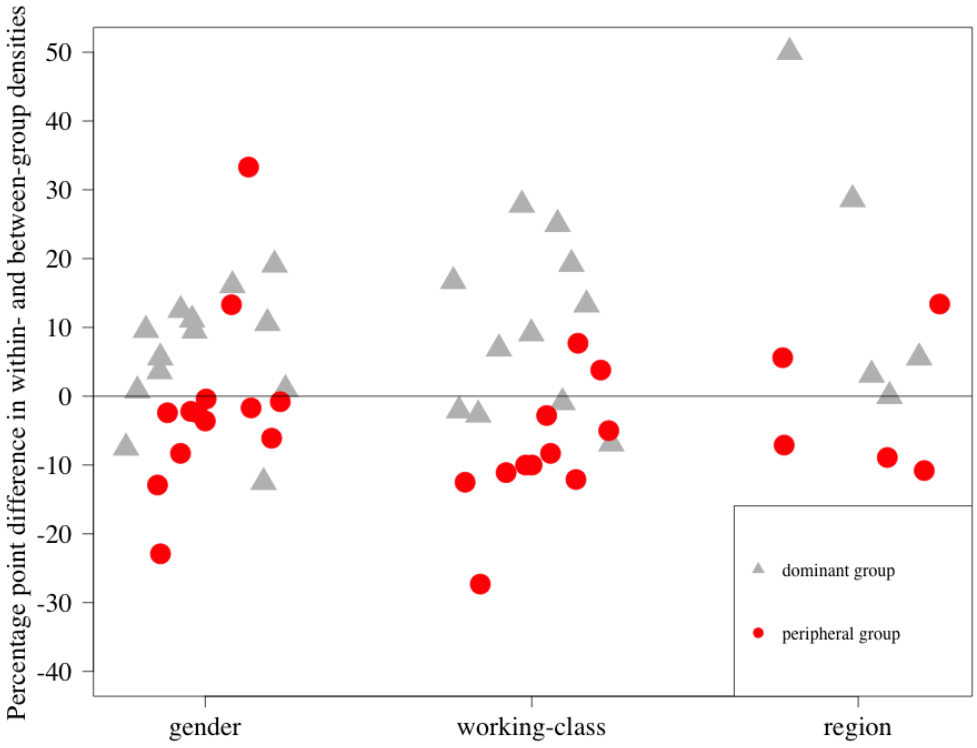


Note: Marginal effect and 95% confidence intervals are based on the hierarchical models presented in Table E.1 in the Appendix E. The shaded areas demarcate the location of the groups in which CEE participants were engaging with participants from older EU-member states (WE, SE, or both). See also Figure E.2 (Appendix E) for the results on the simulation of the conditional effects.

In sum, results indicate that the magnitude of the decrease in the representation of women, CEE members and working-class participants is substantial at the level of contribution. While not so pronounced in terms of frequency of speaking, traditionally disadvantaged groups have a strong tendency to keep their statements short. Yet finding that these groups nevertheless managed to get themselves heard would render previous results less worrisome. As can be gleaned from Table 3.2, participants receive consideration from others in general accordance with their contribution to the discussion: for all three groups, the decrease in consideration lies in-between the (less pronounced) decrease in frequency and the (more pronounced) decrease in speaking time. At first sight, this indicates that no additional mechanism of exclusion was present in relation to consideration of arguments. Yet we may also want to know whether arguments of the more peripheral groups have been fairly considered by members of the dominant group. Discovering that views of marginalized group members only found appropriate consideration among members of the peripheral groups would reveal the

previous optimism as too hasty. Therefore, we turn to the question of homophily. Homophily analysis explores whether individuals' views were more likely to be considered by similar others. While the absence of homophily would suggest that the claims of peripheral groups reasonably entered the consideration process of members of the dominating groups, evidence of homophily might imply that exclusionary mechanisms were at work. I test for homophily by comparing network densities of the consideration networks between peripheral and dominant groups (see Lazer et al., 2010). Network densities are calculated as the total number of existing relations (here: references) over the total number of possible relations. Densities can take in any number between 0 (here: indicating that no one's view was referenced) and 1 (indicating that everyone's view was referred to by all members of the group). Figure 3.3 displays the percentage point differences in within- and between-group densities. Positive numbers indicate a tendency for homophily (more references were made to in-group members), while numbers below zero indicate a tendency toward heterophily (more references were made to out-group members). While the triangles display the differences in densities for the more dominant cohorts (separately, for each small group discussion), the circles show the same from the perspective of the marginalized groups.

Figure 3.3: Homophily Effects for Members of the Dominant and Peripheral Groups

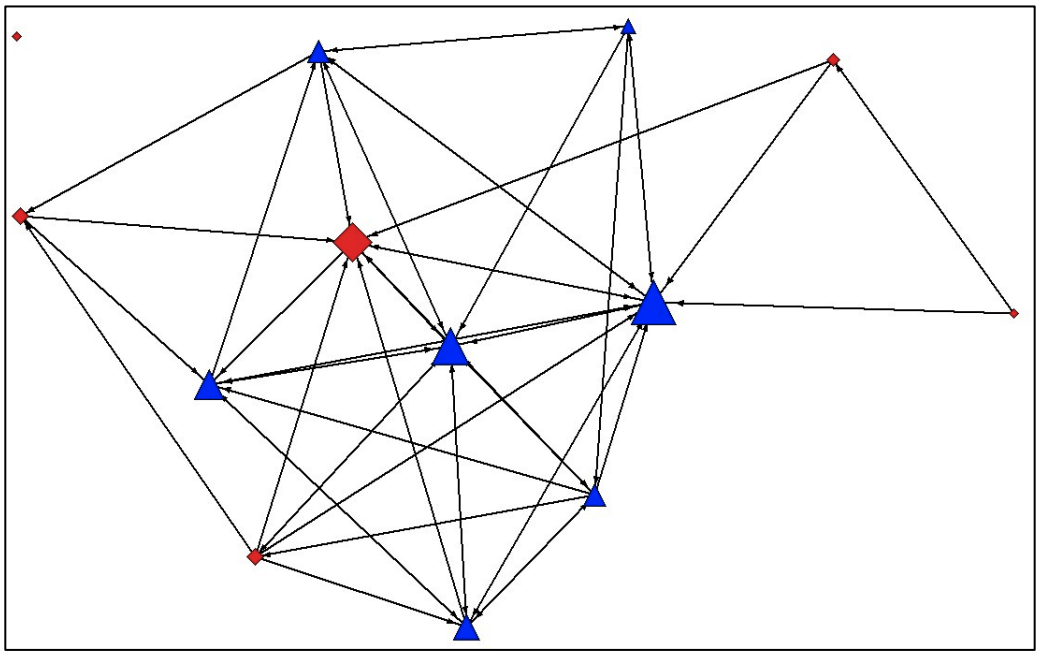


Note: For the corresponding entries to the Figure, see Tables E.3-E.5 in Appendix E.

Figure 3.3 reveals an overall tendency towards homophily for men, non-working-class participants and citizens of older EU-countries: the dominating groups in terms of

contribution tend to consider views from their peers more often than views of their more peripheral discussion partners. Yet the extent to which they do that varies greatly from group to group. For example, while in three groups, participants from older EU-countries were equally likely to interact with participants from CEE countries as with their peers, another group shows a 50 percentage point difference in favor of referencing their dominant peers. As an illustration, Figure 3.4 graphs the consideration network of the respective group.

Figure 3.4: Network of Consideration (Group 9)



Note: Attributes display country affiliation. Diamonds are used for CEE participants (Polish), triangles for WE participants (Germans and Austrian)). The size of the nodes is given by the in-degree centrality of an actor (amount of references received from co-discussants). The arrows indicate the direction of the reference. Isolates (top left corner) indicate participants that contributed to the discussion but neither made a civil reference to someone nor received one. Graphs were created using NetDraw (Borgatti et al., 2002).

Turning to the marginalized groups (circles), the reverse picture is on display. In the large majority of the group discussions, women, working-class participants and members from new EU-countries showed higher consideration densities towards members of their out-group. For some reason, members of peripheral groups deemed insights of the dominating crowd more pivotal than perceptions of members of their own groups. Yet there are two notable exceptions for gender: in two comparatively homogeneous groups (group 3, WE members only; group 16, CEE members only), women received more references from both male and female participants.

3.5 Discussion and Implications

At this point, we may only speculate on the reasons for the patterns of dominance detected in the foregoing analyses. One may argue that, in terms of epistemic quality, the contributions of members of peripheral groups may not have lived up to the contributions provided by the dominating speakers. Recognizing some actors to be the more eloquent advocates may have intimidated other actors from contributing and it may explain why the formers' views were more likely to find consideration. Yet we demonstrate in Chapter 2 that the deliberative competence of women and members of new EU-countries who participated in Europolis does not significantly rank behind their dominating counterparts (see also Himmelroos, 2011).[34] The results are different for working-class members: their contributions were of lower deliberative quality than those of higher class participants. According to the previous analysis, working-class participants are also the most marginalized group in terms of consideration. Yet, as discussed above, deliberative theory would be unwilling to *a priori* exclude a particular group from influencing the deliberative process, particularly if the group is likely to add a relevant social perspective to the discussion (Bohman, 2006; Brown, 2006).[35] By discussing third country migration, working-class participants could have offered a new perspective on blue-collar immigrants that could have been deemed worthy of consideration. The same accounts for the new CEE members and their potentially different experiences with migration. On the other hand, it is hard to understand what particular perspective men, as a group, could have introduced to the discussion on third country migration that would justify their dominant position in the discussion.

Yet to conclude that the findings are the result of purposive exclusion by the dominating group may be too hasty, given the results of the homophily analysis and an extension thereof: in 8 out of 13 groups (11, respectively), women and working-class participants received, in absolute terms, more reference from the central actors than from members of the peripheral groups.[36] Thus, the fact that marginalized groups received somewhat less consideration compared to their dominating counterparts is not necessarily due to a lack of interest from the side of the central groups. Rather, marginalized actors seem to contribute to the preservation

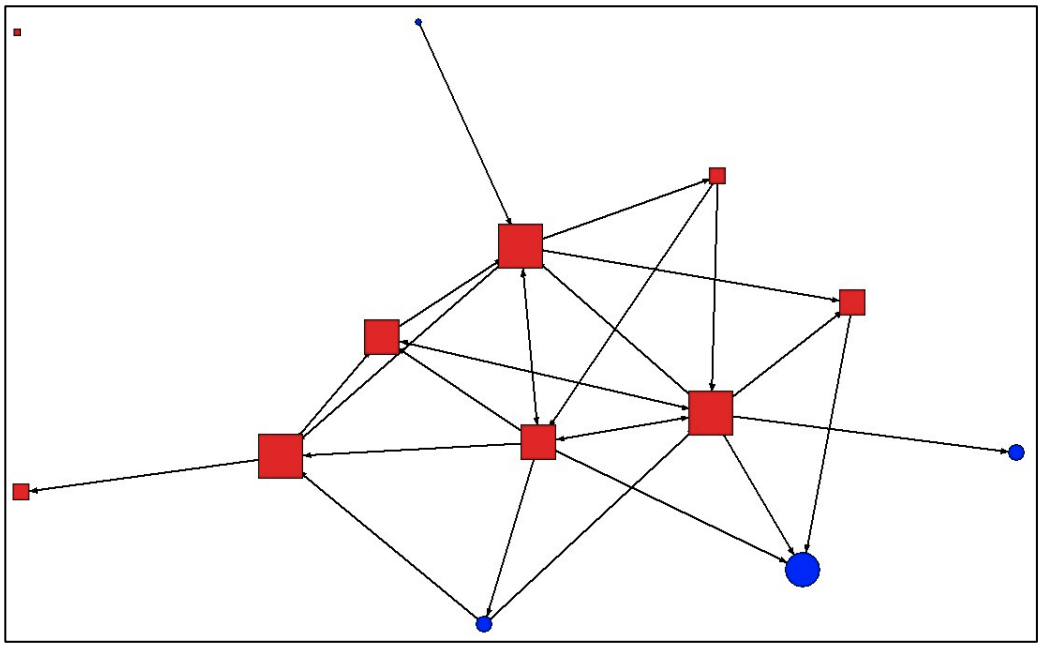
[34] However, it should be noted that in Europolis, the correlation between latent deliberative quality and contribution in terms of frequency and speaking-time is moderate (see Chapter 2, section 2.6). This accounts as well (to a slightly lesser degree) to consideration ($r = 0.616$).

[35] A social perspective does not necessarily stand for shared interests or preferences. Rather, a social perspective consists of „shared experiences (...) which tend to give rise to shared questions and concerns" (Brown, 2006: 218).

[36] The same is true for 3 out of the 5 heterogeneous groups that included more than one member of a new EU-country. See Tables E.3-E.5 in Appendix E.

of their peripheral position by considering the arguments of the dominating speakers as more valuable. Figure 3.5 illustrates such a situation in group 1.

Figure 3.5: Network of Consideration (Group 1)



Note: Node attributes are shown for gender (men as squares, women as circles). The size of the nodes is given by the in-degree centrality of an actor (amount of references received from co-discussants). The arrows indicate the direction of the reference. Isolates (top left corner) indicate participants that contributed to the discussion but neither made a civil reference to someone nor received one. Graphs were created using NetDraw (Borgatti et al., 2002).

Are these findings concerning in light of what we know about inequalities in other deliberative settings? A tentative answer is both yes and no. Compared to real-world deliberation in town meetings (Bryan, 1999; Mansbridge, 1983), inequalities in contribution are of minor concern in Europolis. Although this suggests that artificially created (mini-)public spaces are able to attenuate common patterns of social dominance, one still wants to know whether they manage to do so to their best ability. Compared to other settings that use similar procedural safeguards, the following pattern emerges: while contribution biases for socio-economic characteristics have now been detected repeatedly in face-to-face deliberative setting (Hansen, 2010; Himmelroos, 2011), they have proved to be almost completely absent in online DP (Siu, 2009; Kim et al., 2010; see Price, 2009). Keeping the latter in mind, I refrain from concluding that remaining discrepancies in contribution patterns between social groups in Europolis are solely the result of different speaking styles and that patterns of social dominance have been completely absent.

Finding that consideration patterns are largely a reflection of the patterns for frequency of contribution recalls Sanders' (1997) meta-analysis of jury studies. Sanders (1997: 365) concludes that whether an argument prevails during deliberation is a matter of whether the argument has a "talkative promoter". A promising (though tentative) conclusion would be that by enhancing contribution equality, consideration biases could be attenuated. On the other hand, offering "overspeaking" as a panacea to recognition may leave us with an uneasy feeling: a participant that only contributed once may offer a novel perspective or introduce critical remarks, while a participant that "overspeaks" may add little to the discussion if he only repeats the same arguments several times. From a practical point of view, one may also question how participation equality could be boosted further, given that DP already places so much emphasis on providing a favorable environment for free and equal participation. Nevertheless, facilitator intervention varied across Europolis group discussions (Gerber, 2011; see also Bächtiger and Gerber, 2012). A systematic evaluation of facilitators' influence on fair and equal participation is, however, beyond the scope of this chapter. What is more, facilitator interventions may be more appropriately studied in experimental settings, where different facilitator styles are purposely introduced as a treatment variable (see Reykowski, 2006; Schneiderhan and Khan, 2008). Yet a deliberative theory of facilitation has hitherto been missing (but see Moore, 2012). Whether and how facilitators manage to establish a fair and equal discussion environment are important issues for future research.

3.6 Conclusion

In this article, I have proposed a pathway to evaluate equal participation during small group discussions in deliberative polling (DP). The main motivation for doing so is that informal impediments to participation may remain even after everyone is formally empowered to take part (e.g., Fraser, 1992; Sanders, 1997; Young, 2002). By accounting for the intrinsically relational aspect of interpersonal communication, the paper adds to the small but growing literature on deliberation that voices reservations about taking the willingness to engage with others claims for granted (e.g., Burkhalter, 2002; Dobson 2010, 2012; Stromer-Galley, 2007; Pedrini et al., 2013). Methodologically, the article responds to Mendelberg (2005) who argues for taking the deliberative context more seriously. As a consequence, multilevel modeling and social network analysis have been proposed as suitable tools of analysis.

In terms of contribution, the findings reveal that particularly women, members of new EU-countries and working-class participants spoke significantly and substantially less than their counterparts. Most importantly, results suggest a contextual effect on contribution: with

increasing diversity of national backgrounds, women and new EU-members tend to give the floor to men and members of older EU-states. Yet the effects are more attenuated for the absolute amount of speaking time, while they are nuanced for the number of speeches made by individuals. When it comes to patterns of consideration, the article is far from concluding that the dominating groups purposely excluded the views of more peripheral actors. Rather, results suggest that the degree to which a speaker receives consideration from others largely depends on the number of speeches he or she contributes to the discussion. Thus, except for the working-class participants whose exclusion from the discussion is consistent and substantial, lacking consideration was not a serious issue for traditionally marginalized groups. Moreover, finding that receiving consideration is strongly bound to one's presence in terms of utterances gives rise to some concerns. Deliberative theory would meet an approach that offers "overspeaking" as a panacea to recognition with reticence.

The article is limited in the scope and generalizability of its findings. First of all, the sample used in this study was small and results await confirmation from further research. Furthermore, the definition of equality offered in this study is strongly bound to the setting under examination. Recalling that the imperative in DP is learning and understanding rather than decision-making (Fung, 2003; see also Olsen and Trenz, 2013), a deliberative process has to make sure that the relevant reasons are aired and appropriately weighed in order for citizens to develop more informed and considered opinions (Mansbridge et al., 2012). In a setting where the aim is decision-making, one might also want to place emphasis on the nature of the deliberative outcome, for example in terms of inclusiveness, justice, or fairness (e.g., Gutmann and Thompson, 2002; Rawls, 1971; Sanders, 1997). In addition, the article is restricted to the question of whether participants actively communicated their arguments and on the uptake of these arguments. In doing so, the nature and consequences of participants' contributions are not to be denied. Rather, they are the topic of concurrent work (see also Chapter 4). The principle aim of the chapter is to call for alternative and multiple evaluative yardsticks of equal participation in deliberative mini-publics.

Chapter 4

DELIBERATIVE AND NON-DELIBERATIVE PERSUASION*

Marlène Gerber, André Bächtiger, Irena Fiket, Marco Steenbergen, Jürg Steiner

* We would like to thank participants at the Special Issue workshop at the University of Stanford, December 2012, as well as participants at the research seminar of the Social Science Research Institute, Abo Akademi (Finland), and participants at the PhD colloquium of the Department of Political Science, University of Turku (Finland), for their valuable comments and insights.

A revised version of this paper has been published 2014 in *European Union Politics* 15(3): 410–429.

4.1 Introduction

To date, the success of deliberative polls (DPs) has mainly been evaluated by assessing whether participants change their minds after deliberation (e.g. Luskin et al., 2002; Fishkin, 2009; Farrar et al., 2010). Opinion change is crucial for the Madisonian approach to deliberation that James Fishkin and associates are strongly wedded to (Morrell, 2010). Madisonian deliberation focuses on engaging citizens (or legislators) in reflective consideration of their own views. Since people are not all knowing, being confronted with new information and hearing other opinions should be conducive to opinion change. From a normative vantage point, however, post-deliberative opinions may not be valid unless we ascertain that they are the results of "mechanisms specified in the normative theories" (Neblo, 2007a: 1). This is also recognized by Fishkin (2010: 70) who acknowledges that "the question for the DP is whether or not the moves are based on reasoned argument with good information considered under good conditions". While Fishkin and associates have already investigated the link between information gain and opinion change (e.g., Luskin et al., 2002), very few scholars have taken the trouble to investigate whether and to what extent post-deliberative opinion is linked to the deliberative quality of the discussion process (exceptions are e.g., Siu, 2009; Sanders, 2012).

Ignoring the deliberative process as a potential driver for post-deliberative opinion is a major deficiency, though. First, most deliberative theorists put a strong prime on the actual process of deliberation and the quality of arguments presented during that process. As mentioned above, opinion change should occur after the 'give and take' of 'good' arguments. However, the claim that deliberation is conducive to enlightened opinions stands in stark contrast to previous evidence from public opinion research and cognitive psychology. Lupia (2002), for example, lists various cognitive constraints that hinder listeners to properly engage with a speaker's message. Consequently, Lupia reproaches deliberative theorists for pursuing an "unreliable folk theory of learning" (p. 149). Similarly, public opinion researchers have hinted at framing effects, i.e. the danger for citizens to change their opinions in direction of the more accessible (i.e. more recently or frequently used) frames (see Druckman, 2001). In order to answer the question whether deliberation actually lives up to its promise, we should attempt to grasp its influence on opinions and specify to what extent it can be separated from other, non-deliberative opinion formation processes. Second, DPs consist of a package of interventions, out of which only the group discussions can be considered truly deliberative. Other interventions, such as the provision of balanced briefing material and the discussions

with experts and/or politicians are basically means to provide participants with additional information (Mutz, 2006: 59). Some deliberative theorists have even claimed that the discussion component may be less meaningful for opinion change than the information phase and the internal-reflective process in participants' heads prior to discussion (what Goodin (2000) calls "deliberation within"). Therefore, we need to better determine the exact causal weight of the deliberative process in the small discussion groups on post-deliberative opinions (e.g., Price and Neijens, 1998; Schneiderhan and Khan, 2008).

The goal of this article is to open up the black box and explore deliberation's influence on post-deliberative opinion formation. In concrete, we ask whether and to what extent deliberative quality in the discussion process of DPs is linked to individual post-deliberative opinions. Theoretically, we combine the formal quality of argumentation - which we dub "justification rationality" - with a focus on the content and direction of argumentation and link it to opinion change. We call this "deliberative persuasion". We directly contrast our conception of deliberative persuasion to non-deliberative persuasion, where the repetition of positions rather than the formal quality of arguments matters for opinion formation. Empirically, we focus on Europolis, whereby we analyzed the discussions on third country migration in a subsample of 13 groups.[37]

The article proceeds as follows. First, we discuss the mechanisms of opinion change as specified in public opinion research and deliberative theory. Next, we present our approach of deliberative persuasion, followed by a description of the data, measures, and the statistical method. The next section presents the results, followed by a discussion of the implications of our results. The last section concludes.

4.2 Two Perspectives on Public Opinion Formation

Both deliberation and public opinion researchers agree that in general, citizens lack information about political matters (e.g., Converse, 1964; Delli Carpini and Keeter, 1996; Ackerman and Fishkin, 2002). Apart from this agreement, the two research traditions have worked on polar tracks for a long time. While public opinion researchers explored the potential for low-information rationality (e.g., Lupia and McCubbins, 1998), deliberative researchers argued that deliberation helps to inform public opinion. In the following, we will first give a brief overview how opinion formation processes are perceived by public opinion

[37] With regard to persuasion, we understand the successful attempt to convince someone of one's opinion on a matter. As such, the definition of persuasion *per se* is value free (see e.g., Lupia and McCubbins, 1998). In the context of our study, persuasion can rely on deliberative or non-deliberative mechanisms.

research and deliberative theory. Then, we connect with recent attempts that started to combine the two perspectives and explore whether and how deliberation actually lives up to its promises.

4.2.1 Public Opinion Research and Opinion Change

Classic research in public opinion formation generally viewed opinion changes as problematic. According to Converse (1964), most citizens are incapable of making valuable contributions to public debates given their ignorance about the substance of politics and their inconsistency of opinions on public issues. Consequently, opinion stability – and not opinion change - was seen as a sign of 'considerateness'. However, public-opinion research in the post-Converse era challenged these assumptions on theoretical and empirical grounds (Bartle, 2000). Zaller and Feldman (1992) argued that the instability of attitudes is the result of more complex information processing related to factors such as issue salience, rhetoric of the political elite, political awareness and individual predispositions rather than a product of non-attitudes. In other words, complex information processing can lead to attitudinal ambivalence (Zaller, 1992; see also Lavine et al., 2012). In the following, the complexity of opinion formation (or, change) was further explained by issue framing in media and elite discourse (Alvarez and Brehm, 2002; Druckman, 2001). The argument is that frames, i.e. the selection and presentation of information, shape how citizens think about an issue. Framing studies showed that exposing people to an argument for a given position has the power to push an individual toward this position (e.g., Zaller, 1992; Nelson et al., 1997). While classic studies of framing effects only exposed people to one particular frame, recent studies started to incorporate competing frames, making such experiments more adapted to political reality (e.g., Sniderman and Thierault, 2004; Brewer and Gross, 2005; Chong and Druckman, 2007). This development is associated with a paradigmatic shift in how citizens are viewed to process information. As Druckman (2001: 241) holds: "If emphasis framing effects work through a more deliberate process, we would expect systematic limits to political manipulation: people will not follow any frame that they recently and/or frequently hear". By demonstrating that balanced discussion has the potential to limit framing effects, opinion researchers began to 'warm up' with the idea that deliberation could have beneficial effects for opinion formation and change (e.g., Druckman and Nelson, 2003; Price et al., 2005; but see Lupia, 2002).

4.2.2 Deliberation and Opinion Change

Frames are unavoidable in all types of communication, even in deliberation (Price and Neijens, 1998; Calvert and Warren, 2012). However, problematic frames may be mitigated in deliberative settings such as DP, where a wide range of different frames (or, perspectives) on a particular issue is deployed and where facilitator intervention is supposed to establish a balanced and civil discussion environment. In such settings, people have the opportunity to "choose their own 'terms of the debate'" (Price and Neijens, 1998: 160), which may make them less prone to framing effects (Hansen, 2007; Calvert and Warren, 2012). The idea of transformation of opinions and preferences is a defining characteristic of deliberative theory. Compared to aggregative visions of democracy, where exogenous preferences are merely cumulated, deliberative theory assumes that preferences are endogenously formed in the political process (e.g., Mansbridge, 1983; Rawls, 1993; Chambers, 2003). Being exposed to "information and arguments that may be dramatically different from what they ordinarily encounter", people should be open to reflect on their own preferences in the light of better arguments (Barabas, 2004: 688). Moreover, being exposed to divergent viewpoints increases the chances that people actually engage in an extensive and deliberate process of reasoning – something which is not always the case in discussions with like-minded or in the context of internal reflection or "deliberation within" (Goodin, 2000; Landemore and Mercier, 2012).

Yet, deliberative theorists are wary of making an *automatic link* between process quality and the propensity of changing minds. It is entirely possible that after serious reflection, participants find out that their original opinions and beliefs are fully valid, and consequently do not change their minds (Barabas, 2004; Dickson et al., 2008). However, when opinion change takes place, deliberative theory would require it to occur "via mechanisms specified in normative theory" (Neblo, 2007a: 1). Thus, individuals should change their minds only if they come to the conclusion that arguments presented by other participants are superior compared to their own (or, other) arguments. To date, this has rarely been subjected to systematic empirical scrutiny. The few existing studies differ widely in their conceptual and measurement approaches as well as in their findings. Two experimental studies found fairly optimistic results suggesting that deliberation rather than mere discussion is conducive to (strong) opinion changes at the group level (Barabas, 2004; Schneiderhan and Khan, 2008). Focusing on DP, Siu (2009) found some evidence that the substantive content of arguments presented in discussion affects the direction of opinion shifts at the individual level. In addition, Siu found that the effect of the substantive content becomes stronger when claims were backed by reasons. Two other studies did not find any significant relationship between

deliberative quality and post-deliberative opinions (Baccaro et al., 2011; Sanders, 2012). While the study by Baccaro et al. (2011) employs an external evaluation of deliberative quality (based on the *Discourse Quality Index*), Sanders (2012) employed a perception-based measure of deliberative quality. In the following, we will expand on these attempts by proposing an approach of deliberative and non-deliberative persuasion.

4.3 Mechanisms of Opinion Change

4.3.1 Mechanisms of Deliberative Persuasion

All deliberative theorists have emphasized the crucial importance of robust reasoning – or, justification rationality (e.g., Cohen, 1989; Habermas, 1991; Rawls, 1993). As Chambers (1996: 99) succinctly puts it, "only rational, that is, argumentative, convincing is allowed to take place". Justification rationality not only makes a speech accessible to rational critique, it also increases the chances of persuading others on the basis of logical coherence. In line with an "argumentative theory of reasoning" (Landemore and Mercier, 2012), we argue that opinion change is normatively desirable when it is based on proper reasoning, i.e., when positions are backed by robust reasons. Our conception of deliberative persuasion starts from the premise that diverse pro- and contra-positions are presented during discussion. Deliberative persuasion then means that specific positions are coupled with well-justified premises. By well-justified premises, we expect, as a minimum, that a complete linkage is made between a premise and a conclusion. Neblo (2007a) calls this "decent rationales". We also consider more elaborated linkages between a premise and a conclusion, whereby participants explain in more detail why they support or reject a specific position or proposal. Neblo (2007a) calls this "sophisticated rationales".[38] All else equal, the expected (and desired) result of deliberative persuasion is that participants shift their opinions in direction of those positions that are embedded in decent or sophisticated rationales. This satisfies the deliberative criterion of a reason-driven pathway to opinion formation.

We may, however, encounter situations where there is a *balance* between decent or sophisticated pro- and contra-rationales. Actually, Fishkin (2009) considers "argumentative balance" as a key feature of a desirable deliberative process. Only when participants have considered arguments for different sides of an issue can we speak of proper deliberation. Indeed, the absence of decent or sophisticated rationales for one side may have nothing to do

[38] Yet, in addition to the formal structure of argumentation, Neblo's (2007a) assessment of decent and sophisticated rationales incorporates basic substantive judgments about the quality of the rationales.

with the superiority of arguments for that specific position but simply with the inability of participants to forward rationales for the 'other side'. Put differently, not every side of a controversial issue may have vociferous and rhetorically gifted advocates, a problem which is amplified in citizen deliberation where participants tend to possess highly varying levels of deliberative ability (see Chapter 2). Therefore, the presence of argumentative balance is generally a welcome fact, even though its consequences are not yet fully explored. On the one hand, argumentative balance may lead to a range of opinion shifts, a fact that is not fully recognized by most deliberative scholars. Clearly, after listening to different positions, participants may weigh respective reasons and find one side more persuasive than the other. But "argumentative balance" may also be conducive to a depolarizing trend, i.e., participants with more extreme positions move towards middle positions. On the other hand, "argumentative balance" may also lead to opinion stability: if participants in deliberation realize that there are 'good' arguments for both sides, they may just stick to their original positions. Finally, it may be that pure balance is not conducive to any opinion change, as Zaller's (1992) model would predict.

A final comment is in order. We are fully aware that high justification rationality may not equal argumentative superiority. For example, a clever speaker might use unnecessarily elaborate arguments precisely to cover weaknesses in the main argument, overwhelming his audience with evidence and complex argumentation. Yet, as Landemore and Mercier (2012: 920) argue, a theory of argumentative reasoning is not detached from truth seeking. Since one function of reasoning is the evaluation of arguments, deliberators have a strong incentive to present sound arguments. If a speaker nevertheless sets up a well-phrased construct of lies, truth seeking is most effectively advanced through falsification of the claim by means of rational argumentation. Moreover, the non-strategic setting of DP (where participants make no consequential decisions at the end) may considerably reduce the incentive to depart from truth in order to unconditionally impose one's (narrow) interests on the group. In this sense, DPs are viewed as "safe havens", that are "insulated from certain negative or distorting effects", as we typically find them in the broader public and political sphere (Chambers, 2004: 400). In addition, one might criticize that our definition of deliberative persuasion reduces deliberative quality to purely formal properties of the force of the better argument (see Neblo 2007b). We agree that the conception of deliberative persuasion that we advance in this paper is a narrow one. However, to provide a comprehensive definition of deliberative persuasion would be beyond the scope of this paper. Future research may want to broaden the conception of deliberative persuasion and find a way to incorporate substantive criteria of deliberative

quality. Yet we defend our approach in relation to the case under examination: in Europolis, justification rationality is not only part of a latent construct of deliberative quality at the level of the speakers (see Chapter 2), the construct can also be sensed at the level of the (most rational) speeches. In general, the speeches with the highest justification rationality are characterized by public spiritedness and the complete absence of disrespectful utterances towards the relevant social group.

4.3.2 Mechanisms of Non-Deliberative Persuasion

Ideal deliberation should be immune against non-deliberative mechanism of persuasion (such as threats, coercion or other kinds of repression and inequality), i.e. mechanisms that are unrelated to the merits of an argument (Chambers, 1996: 100; see also Neblo, 2007a; Mutz, 2008). However, social and cognitive psychology as well as rational choice theory has listed various cognitive constraints that may hamper deliberatively induced opinion change (see Lupia, 2002 for a review and Neblo, 2007a for an empirical test). Here, we focus only on one non-deliberative mechanism, namely the influence of a dominant position. In order for deliberation to have epistemic value, we need to make sure that the evaluation of arguments also occurs independently of the *size* of the group offering them (Estlund, 1997: 191; see also Myers, 2012). Majorities might not always be correct. If latent social pressures make people change their minds in direction of the majority's viewpoint, "social muscle, not persuasive [deliberative] argumentation, carries the day" (Mendelberg, 2002: 159). Yet, one may argue that mechanisms of social comparison are less frequent in DP, where participants are not forced to reveal their post-deliberative preferences to the rest of the group. Yet, skewed positioning in a specific direction may still exert a non-deliberative persuasion effect. Similarly, research in psychology has shown that moderate levels of repetition of arguments and statements may push uninformed individuals in this direction, by creating a so-called "illusion of truth*"* or validity effect (see, e.g., Arkes and Blumer, 1984; Moons et al., 2009; see also Chong and Druckman, 2013). Therefore, we complement our analysis by assessing whether the dominant position affected the post-deliberative opinions of Europolis participants.

4.4 Data and Measurement

Empirically, we analyzed the discussions on third country migration in a subsample of 13 Europolis discussion groups by employing OLS regression with clustered standard errors. We refrained from analyzing all 25 small discussion groups, since analyzing deliberative

processes empirically is a highly demanding and time-consuming affair. Although our subsample is limited in national diversity compared to the full sample of participants, it is comparable to the rest of the sample with regard to pre- and post-deliberative opinions on our questions of interest (see Appendix D).

4.4.1 Dependent Variables

In this study, we concentrate on individual level opinion change rather than on group level outcomes, such as the net aggregate opinion shift of a group. The point is that opinion changes at the individual level may occur in the absence of a net shift at the aggregate. Moreover, we do not create an indicator of immigration preferences. Although the participants were required to answer a batch of questions on preferences and attitudes toward third country immigration, many of those questions were rather specific and were not intensely discussed in the groups. Therefore, we focus on two preference items in connection with the immigration topic: "regularizing irregular immigrants" and "European involvement in immigration affairs". These two items formed the core of the Europolis immigration discussions and were rather intensively discussed within the groups, which is essential when exploring the consequences of different deliberative abilities (see Stromer-Galley, 2007: 6). While the item "European involvement" represents a fairly pragmatic issue, the item "regularizing irregular immigrants" has a strong humanitarian dimension (Kymlicka, 2000). Irregular immigration was measured on an 11-point scale whereby participants had to rank their policy choice between wholehearted agreement with return (disagreement with regularization) (0) and wholehearted agreement with regularization (disagreement with return) (10). *European involvement in immigration affairs* was measured on an 11-point scale whereby participants had to rank the level at which third country immigration should be regulated by the European Union or by individual member states. While (0) means "entirely at the individual member states", (10) denotes that the decision should be left completely to the European Union. We evaluate both items at the end of the Europolis event (T3), controlling for opinions at T2 (immediately before the discussion started). This strategy to assess opinion change has been used elsewhere (Neblo, 2007a; Sanders, 2012). We refrain from measuring opinion change as a difference between pre- and post-deliberative opinion due to "ceiling effects" (Luskin, 2002; see also Luskin et al., 2002). For example, individuals already having a progressive position at T2, i.e. fully in favor with the regularization of irregular migrants, cannot become more progressive at T3. If the best argument in terms of justification rationality was in favor of regularization of migration, such individuals can only

keep their initial position (see Barabas, 2004). In a model of deliberatively-induced opinion change, such individuals would receive a value of 0 (indicating no change), putting them at the same level with individuals that did not change their minds at all, but could have done so because their initial opinions were less favorable toward regularization.

4.4.2 Predictor Variables: Deliberative and Non-Deliberative Persuasion

To obtain a measure for deliberative persuasion, two external coders identified all relevant statements that Europolis participants offered on policy options for irregular immigration and European involvement in immigration affairs. In addition, the coders also assessed the direction of arguments by looking at whether the speakers expressed acceptance or refusal of the specific policy option. The coding categories are largely based on options that were presented in the briefing material.[39]

We also assessed the justification rationality of these statements on the basis of the Discourse Quality Index (DQI) developed by Steenbergen et al. (2003). The DQI captures justification rationality under the rubric of *level of justification* and is a measure to assess whether and to what extent statements, or speech acts, have been backed with reasons. Basically, the DQI distinguishes four levels of justification: (0) no justification provided for a statement; (1) inferior justification where the linkage between reasons and conclusion is incomplete (i.e., by merely providing illustrations without linking them to the conclusion);[40] (2) qualified justification where a single but complete linkage between reasons and conclusion is made; (3) sophisticated justification where a problem is examined in-depth and/or from various viewpoints.

In line with Neblo (2007a: 20), we see the provision of argumentative rationales as a group source and transform our speech level data into a group level variable of "deliberative persuasion". For each discussion group, we first focused on progressive statements, i.e. statements in favor of regularization of irregular immigrants or statements against sending irregular immigrants back, and contrasted them with conservative statements, i.e. statements favoring sending irregular migrants back or statements opposing regularization. We then

[39] This coding was part of a larger coding project where we coded all statements that participants uttered in the immigration discussions. In total, we coded acceptance or refusal of more than 20 policy options and assessed whether actors considered a certain aspect of immigration as problematic or beneficial.

[40] We are far from considering stories, or illustrations, as inferior modes of communication. In Chapter 2, we present an updated version of the DQI that accounts for storytelling's importance by capturing its presence in a separate category. Moreover we even found story-telling to be a partial complement to high justification rationality. However, in cases where illustrations are used without linking them to the conclusion, the statements lack justification rationality.

applied the same procedure for statements on the item of European involvement in immigration affairs. Here, we identified Europeanist statements, i.e. statements in favor of a European immigration policy or statements against country specific solutions, and contrasted them to nationalist statements, i.e. statements against a European solution or statements favoring country specific solutions. Having identified progressive (Europeanist) vs. conservative (nationalist) arguments, we then assessed the level of justification for the various "argumentative camps". For each "argumentative camp", we identified the statement with the highest level of justification and compared it to the statement with the highest level of justification of the opposite camp. In each group and for both items, the following three situations might arise: (1) *Progressive/Europeanist position with higher justification rationality*: At least one of the progressive/Europeanist statements had a higher level of justification than any conservative/nationalist statement presented during group discussion. (2) *Argumentative balance*: the level of justification of progressive/Europeanist statements was equal to the level of conservative/nationalist statements. (3) *Conservative/nationalist position with higher justification rationality*: At least one of the conservative/nationalist statements had a higher level of justification indicator than any progressive/Europeanist statement presented during group discussion.

We provide a concrete coding example in Appendix G.[41] In the end, we are left with three dummy variables at the group level indicating whether situation 1, 2, or 3 applies.[42] We opted for running three separate models with three binary variables, since we wanted to model the effect of one particular situation independent of the two other situations.

In order to obtain a measure for non-deliberative persuasion, we created another group level variable, termed "dominant position". This variable ignores justification rationality and only focused on the frequency of specific policy positions over others (see also Siu, 2009 for a similar approach). Our variable divides the total number of arguments in favor of regularization of irregular immigrants (or, in favor of European involvement) by the total number of statements presented on the respective issue (i.e. statements in favor and statements against regularization of irregular immigrants – or, European involvement).

[41] In all groups, justification rationality for the statement ranked the best, mounted at least to a score of 2 (qualified justification). However, on both questions, at least one camp in the large majority of groups offered sophisticated justifications (3) for their statements.

[42] Of course, aggregating the speech acts in this way implies losing information. Yet, this kind of aggregation is exactly what a proper test of deliberative theory requires, namely investigating whether the 'best' argument won – independent of how frequently it has been brought up.

Finally, we report the reliability of our coding. We conducted two inter-coder reliability tests, one for justification rationality, and another one for the content and direction of statements. With regard to justification rationality, three independent research assistants coded a random sample of statements (N=80). The training of the coders consisted only of the instructions contained in the coding scheme (see above) and a brief verbal introduction by the project leader. For level of justification, coder agreement was 65% and Cohen's Kappa reliability statistic (κ) mounted to 0.52. Thereby, we missed the benchmark for substantial agreement (κ=0.60) according to Landis and Koch (1977). Spearman's *r* and the standardized Cronbach's alpha were respectable (*r*=0.78; α=0.92), showing that the existing coding discrepancies resulted from minor disagreements. After the disagreements had been settled, a second reliability test round has been performed before coding the full database. Now, the re-test (N=20) indicates substantial coder agreement with a rater coder agreement of 80% and Kappa of 0.73, and with scale reliability statistics of 0.97 for Cronbach's alpha and 0.94 for Spearman's *r*. With regard to the content and direction of statements, three coders conducted a reliability test by coding 30 randomly chosen statements. A value of 0.67 for Cohen's Kappa indicates satisfactory inter-coder reliability for the content of statements. The direction of statements was never disputed.

4.4.3 Control Variables

We include a number of controls in our models that opinion researchers deem important in the context of opinion formation. We first control for a number of socio-economic and cultural variables: gender (coded 1 for men and 0 for women), participant's age (in years), education (measured as age at which main education ended), working-class (dummy variable) and a categorical variable that allows us to distinguish between participants coming from Central Eastern, Southern or Western European countries. Next, we include a variable that captures an individuals' ideology by asking participants to indicate their ideological orientation on a scale from left (0) to right (10). Further, we introduce a variable for *salience* that indicates the perceived seriousness of the immigration problem on an 11-point scale from (0) ("not at all") to 10 ("most serious problem we face"). We also include a knowledge index that calculates the amount of correct answers to the three knowledge questions on immigration (T2) and an item that captures a person's general interest in politics on an 11-point scale from (0) ("not at all") to 10 ("passionately").

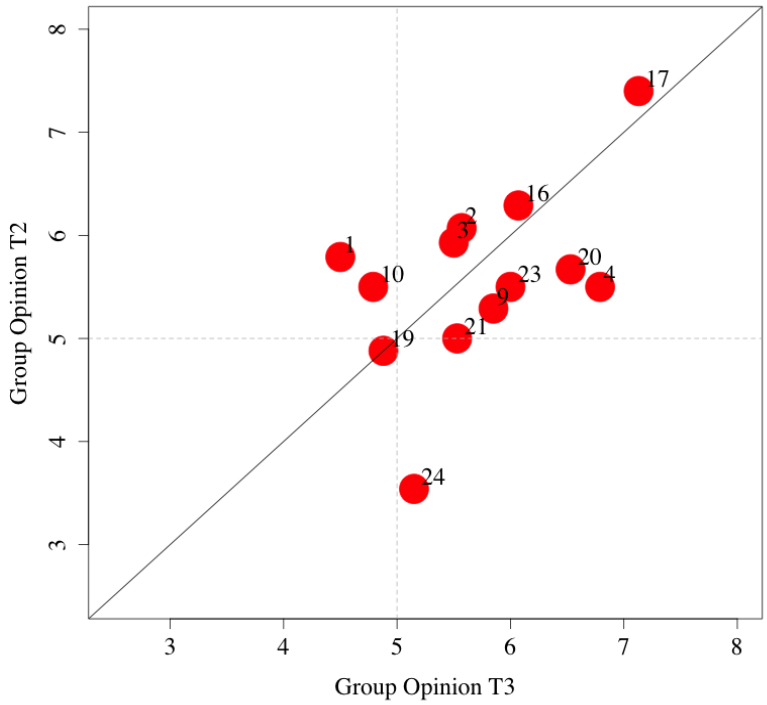
4.4.4 Statistical Analysis

We employ OLS regression to explore whether deliberative persuasion is linked to opinion shifts. All models use clustered standard errors at the level of the groups in order to take into account that errors are potentially correlated within groups. For the question on European involvement in immigration affairs, we limit our sample to ten groups. This is necessary since three groups only offered arguments in favor of Europeanization. Including these three groups in the analysis would preclude us from separating deliberative from non-deliberative persuasion effects. Since the number of groups is small, we only use a restricted number of control variables. We excluded control parameters that showed a very high degree of uncertainty in an initial Bayesian hierarchical linear model incorporating all predictors. For regularizing irregular immigration this concerned age, working-class, interest in politics, and the indicator for country regions; for European involvement in immigration affairs, this concerned working-class, education, and the indicator for country regions (see Appendix F).

4.5 Results

We start with post-deliberative opinions on irregular immigration. Overall, we find no net aggregate shift for this item. At T2, participants of our subsample had an average score of 5.6 on the 11-point scale, indicating a marginal favourability towards regularization (10) compared to sending irregular immigrants back (0). After deliberation at T3, the average score was 5.7 (see Appendix D). Despite the absence of a net aggregate shift, 60.5% of participants shifted their opinions by at least one scale point (and 36.8% by two or more scale points). That the groups display different pattern of opinion shift becomes visible by glancing at the scatterplot in Figure 4.1. While on the aggregate, some groups became more open to regularization (area below solid line), others became more open to sending irregular immigrants back (area above solid line).

Figure 4.1: Aggregated Pre- and Post-Deliberative Opinions on the Item "Regularization of Irregular Immigrants" (by Group)



The next question we would like to address is whether these group-specific differences can be explained by factors that occurred in the small group discussions. Using OLS regression, we now explore whether our mechanisms of deliberative and non-deliberative persuasion contributed to explain post-deliberative opinions (T3). Table 4.1 reports the results. We first explore the effects of deliberative persuasion (Model 1a and 1c). We find that participants in groups where the progressive position entailed a higher level of justification (7 out of 13 groups) were significantly more open to regularization after deliberation (Model 1a). On the other hand, individuals in groups where the conservative position had a higher level of justification (2 out of 13 groups) were significantly less open to regularization after deliberation (Model 1c). This is excellent news for our conception of deliberative persuasion, emphasizing that opinion change takes on deliberative virtue when positions are backed by robust reasons. However, the effect size of deliberative persuasion is rather small. Being exposed to higher quality progressive arguments leads to a 0.813 increase in favor of regularization on the 11-point scale (Model 1a). If the higher quality argument was conservative, the model predicts a 0.693 decrease on the scale (Model 1c). This effect is comparatively small if one takes into account that a three-point difference on the salience index already accounts for (almost) the same effect size. In other words: even though

mechanisms of deliberative persuasion were at work, they were not able to compensate for effects of predispositions such as salience, ideology or knowledge (see also Hansen, 2007).

Table 4.1: Predicting Post-Deliberative Opinions for the Item "Regularization of Irregular Immigrants"

	Opinion T3		
	Model 1a	*Model 1b*	*Model 1c*
Opinion T2	0.341***	0.353***	0.363***
	(0.081)	(0.082)	(0.079)
Men	-0.562^{+}	-0.555^{+}	-0.514^{+}
	(0.285)	(0.295)	(0.286)
Education	-0.031	-0.038	-0.039
	(0.028)	(0.028)	(0.028)
Ideology	-0.175***	-0.176***	-0.183***
	(0.038)	(0.039)	(0.040)
Knowledge T2	0.395*	0.394*	0.367*
	(0.136)	(0.138)	(0.130)
Salience T2	-0.256**	-0.249***	-0.250**
	(0.061)	(0.061)	(0.063)
High justification rationality (progressive)	0.813*		
	(0.324)		
Balanced justification rationality		-0.299	
		(0.394)	
High justification rationality (conservative)			-0.693*
			(0.311)
Dominant position	-0.599	0.090	0.087
	(0.576)	(0.694)	(0.739)
Constant	5.430***	5.650***	5.762***
	(0.772)	(0.792)	(0.862)
Prob > F	0.000	0.000	0.000
Adj. R^2	0.414	0.397	0.405

Note: Robust standard errors (clustered at the level of the groups) are shown in parentheses. N=170, Groups=13. $^{+}$p<.10, *p<.05, **p<.01, ***p<.001. T2 indicates participants' opinions right at the start of the three-day event and T3 right at the end of the three-day event.

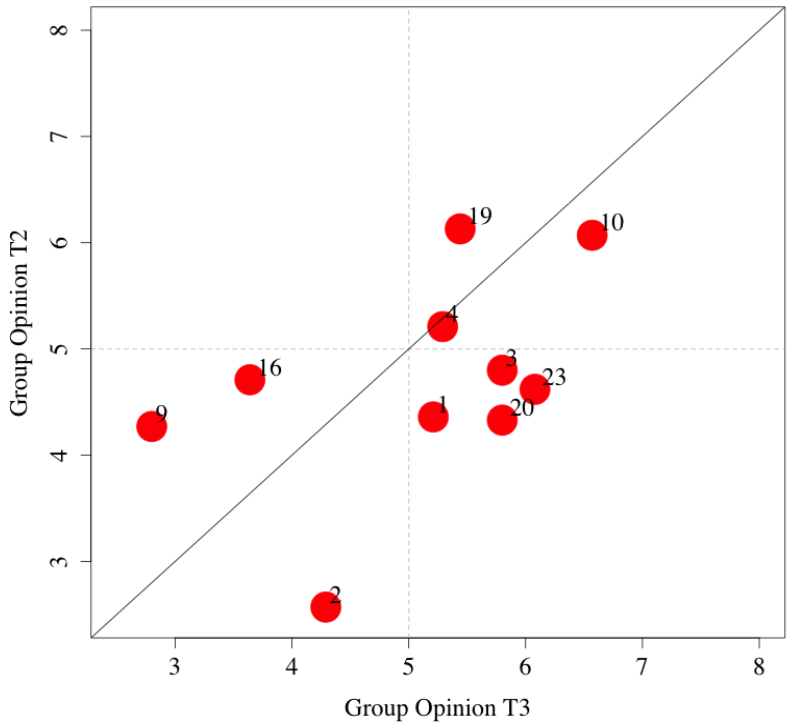
We now turn to Model 1b. When participants were exposed to equal argumentative quality (4 out of 13 groups) there is no significant effect for the direction of opinions. In addition, descriptive results show that participants exposed to equal argumentative quality were not particularly more likely to stick to their original position than participants exposed to unequal argumentative quality. Moreover, polarization was even less frequent in the balanced

groups.[43] In sum, balanced argumentative quality neither led to opinion paralysis nor to confirmation bias, which both is good news for deliberation.

By the same token, participants did not shift their opinions towards the dominant position. The perceived correctness of a position did not depend on how often a position has been brought up. In sum, inherent qualities of a statement, not the mere frequency of a position, influenced post-deliberative opinions.[44]

Let us now turn to the item of European involvement in immigration affairs. The net aggregate shift for this item was small and not statistically significant. Participants in our subsample of 10 groups moved from an average of 4.7 to a slightly more Europeanist position of 5.1. Again, individual level opinion shifts were considerable: A total of 68% of participants shifted their opinions by at least one scale point and 54.2% changed even 2 or more scale points on the 11-point scale. Again, a scatterplot of aggregate pre- and post-deliberative opinions shows that groups differ in their pattern of opinion shift, suggesting that group-specific factors have contributed to shaping opinions in Europolis.

Figure 4.2: Aggregated Pre- and Post-Deliberative Opinions on the Item "European Involvement in Immigration Affairs" (by Group)



[43] Results available on request.

[44] The dominant position in a group was far from always being the position that had the higher justification rationality: the point biserial correlations for our deliberative and non-deliberative group level variables are moderate to low with 0.55 (Model 1a), -0.43 (Model 1b), and -0.21 (Model 1c).

Turning to the OLS regression in Table 4.2, results show that the magnitude of individual-level opinion shifts is so large to that the pre-deliberative opinion on Europeanization does not yield a significant effect on post-deliberative opinion. Table 4.2 shows that there are no statistically reliable effects for deliberative persuasion: reasons in favor or against European involvement did not affect post-deliberative opinions (Models 2a and 2c).[45] By contrast, the main factor driving post-deliberative opinions was dominant positioning: the higher the share of pro-European statements, the more participants were in favor of European involvement at T3.[46]

The models suggest that exposing participants to completely one-sided argumentation either in favor or against Europeanization shifts participants' opinions 4 scale points in either direction. Surely, in DPs, where the stated goal is to expose participants to argumentative balance, completely unbalanced argumentation is a rare occurrence. For the item European involvement in immigration affairs, however, our subsample shows considerable variation in argumentative balance.[47] Thus, dominant positioning is a possibility in DPs with considerable effects on opinion formation. What is more, not only are opinions at T3 driven by the dominant position and cognitive and motivational factors, they are also determined by the socio-economic variables of gender and age. If we understand successful deliberation in terms of increased awareness of others' perspectives (e.g., Mansbridge, 1996b; see Neblo, 2007a), these results give rise to some concern.

At this state of research, we can only speculate about possible reasons why we find a deliberative persuasion effect for one item and a non-deliberative persuasion effect for the other. Obviously, the two items clearly differ in terms of scope, substance and implications, which makes them not comparable to each other. Further research might want to examine whether some issues are more susceptible to deliberative persuasion than others. Given that the best arguments in terms of justification rationality are characterized by the absence of narrow group interests and disrespectful utterances toward migrants, we exclude the possibility that the lack of impact on the Europeanization item may be explained by shallow or inadequate reasoning. To what extent substantive criteria of deliberative quality are

[45] In our reduced sample of 10 groups, 3 groups displayed a quality bias towards Europeanization: 2 groups were biased towards country specific solutions and 5 groups showed balanced argumentative quality.

[46] Again, the most dominant position was by far not always the position with the highest justification rationality. The point biserial correlations for the group level variables are low with 0.21 (Model 2a), -0.03 (Model 2b), and -0.20 (Model 2c).

[47] While the least Europeanist group of our subsample offered 37.5% of relevant statements in favor of European involvement, the most Europeanist group offered 86.7% in favor of European involvement. Remind that three groups were excluded from our model because all statements were favoring European involvement.

supposed to shape the opinion formation process and enter the conception of deliberative perception, however, needs to be addressed in future work.

Table 4.2: Predicting Post-Deliberative Opinions for the Item "European Involvement in Immigration Affairs"

	Opinion T3		
	Model 2a	*Model 2b*	*Model 2c*
Opinion T2	0.139 (0.114)	0.151 (0.115)	0.146 (0.105)
Men	1.021* (0.414)	1.021* (0.410)	1.017* (0.403)
Age (in 10 years)	-0.260* (0.086)	-0.245* (0.076)	-0.250* (0.085)
Ideology	-0.213^{+} (0.100)	-0.204^{+} (0.101)	-0.202^{+} (0.097)
Knowledge T2	0.625** (0.142)	0.612** (0.138)	0.613** (0.137)
Salience T2	0.220* (0.077)	0.224* (0.075)	0.213* (0.069)
High justification rationality (Europeanist)	0.367 (0.339)		
Balanced justification rationality		0.008 (0.485)	
High justification rationality (nationalist)			-0.435 (0.609)
Dominant position	4.249** (1.194)	4.458** (1.258)	4.276** (1.076)
Constant	3.588* (1.191)	3.384* (1.289)	3.692** (1.083)
Prob > F	0.001	0.001	0.001
R2	0.301	0.298	0.302

Note: Robust standard errors (clustered at the level of the groups) are shown in parentheses. N=138, Groups=10. $^{+}$p<.10, *p<.05, **p<.01, ***p<.001. T2 indicates participants' opinions right at the start of the three-day event and T3 right at the end of the three-day event.

4.6 Conclusion

This paper made a first attempt in assessing the influence of deliberative and non-deliberative persuasion mechanisms on individual post-deliberative opinions in the context of Europolis. The findings are mixed and the effects complex. When it comes to the issue of regularizing irregular immigrants, we find evidence that participants moved towards the position with the highest justification rationality rather than towards the dominant position, i.e., the position that was most frequently mentioned during discussion. Yet, this finding does not reflect a general trend: regarding the issue of European involvement in immigration affairs, it is the dominant position and not the formal quality of arguments that drives opinions.

Our article is a modest contribution to the perennial question under which circumstances a statement in deliberative discussions "wins or loses the battles of attention, memory, and choice" (Lupia, 2002: 149). On the one hand, we show that rational argumentation can trump the dominant position, which is excellent news for deliberative theory. On the other hand, we find that mechanisms of non-deliberative persuasion may also be at work. This, in turn, underlines that post-deliberative opinions are not always the result of a careful weighing of competing rational arguments. At this stage of research, however, we cannot say why we find a *deliberative persuasion* effect for one preference item and a *dominant position* effect for the other preference item. Future research will need to clarify under what conditions deliberative and non-deliberative persuasion mechanisms are operative.

Future research will need to broaden the conception of deliberative persuasion. In this article, we did not develop a general or encompassing approach of deliberative persuasion. Our only goal was to investigate to what extent the formal structure of an argument can affect opinions. Yet, several authors have claimed that a good argument cannot exclusively be defined by its formal properties (e.g., Gutmann and Thompson, 2002; Chambers, 2004; see Neblo, 2007b). Mucciaroni and Quirk (2010), for example, have focused on the accuracy and realism of claims. Moreover, one may also wonder whether some arguments should have more leverage than others, regardless of their formal quality. When epistemic questions are at stake, there may be a correct (or, better) answer (at least sometimes), making the formal quality of arguments obsolete. When it comes to humanitarian issues, Neblo (2007b) reminds us that there is root in deliberative theory arguing that deliberation should lead to an expanded sense of community while simultaneously transforming prejudices and stereotypes. Neblo (2007b: 548) dubs this "progressive vanguardism"[48]: "On this understanding, deliberative democracy

[48] Neblo is not an adherent of this position; he only lists it as a possibility how to evaluate deliberative outcomes.

is intrinsically and primarily an emancipatory project with strong substantive content, more or less tracking leftist political concerns". From the perspective of "progressive vanguardism", humanitarian positions should certainly be contrasted with other positions, but the expectation is that humanitarian positions involve a higher level of generalizability and consequently have a higher persuasive leverage. In other words, deliberative persuasion in humanitarian directions should be conducive to opinion shifts in this direction, even in the presence of sophisticated counter-positions. From this perspective, our findings for the legalization of irregular immigrants become mixed: on the one hand, and in line with "progressive vanguardism", the well-argued immigrant-friendly and pro-humanitarian position affects opinion change in immigrant-friendly directions. On the other hand, that position is not the only one with persuasive leverage: robust reasoning in direction of less immigrant-friendly and conservative directions (such as the 'pull'-argument that regularization of irregular immigrants attracts more immigrants) also drives opinion change. "Progressive vanguardists" will be reluctant to interpret this finding as deliberative persuasion. Surely, "progressive vanguardism" is a contestable position, but it may nonetheless provide us with a critical (and provocative) yardstick that future research may not so easily discard.

In conclusion, our research represents a starting point for more systematically linking processes and outcomes of deliberation, a task hitherto largely neglected by empirical research. Our article shows that in a supportive environment for deliberation, argumentative rationality is one mechanism of deliberative persuasion that links the deliberative process to outcomes in a normatively desirable way.

Appendix A

DQI CODING AND THE DEFINITION OF THRESHOLD VALUES

Table A.1: Operationalization and Aggregation of Discourse Quality

	DQI categories	Threshold Value
Justification rationality	(0) *No Justification:* The speaker does not present any argument or only says that X should or should not be done, but no reason is given. (1) *Inferior Justification:* Here a reason Y is given why X should or should not be done, but no linkage is made between X and Y—the inference is incomplete or the argument is merely supported with illustrations. (2) *Qualified Justification:* A linkage is made why one should expect that X contributes to or detracts from Y. A single complete inference already qualifies for code 2. (3) *Sophisticated Justification:* At least two complete justifications are given, either two complete justifications for the same demand or complete justifications for two different demands.	(0) max < 3 (1) max = 3
Common good orientation	(0) Explicit statement concerning constituency or group interests *(own country)* (1) No reference: The speaker does not refer to benefits and costs at all. (1.5) Explicit statement in terms of both, Europe and own country (2) Explicit statement in terms of a conception of the common good in utilitarian or collective terms *(EU, Europe)*. (3) Explicit statement in terms of the difference principle *(solidarity, quality of life, global justice, etc.)*	(0) max < 1.5 (1) max ≥ 1.5
Respect and interactivity	(0) Negative (disrespectful) reference to other participants' arguments. (1) No reference to other participants' arguments. (2) Neutral reference to other participants' arguments. (3) Positive (explicitly respectful) reference to other participants' arguments. (up to three references were coded per statement)	(0) max < 3 (1) max = 3
Respect towards groups	(0) *No Respect*: This code is reserved for speeches in which there are only or predominantly negative statements about the groups. (1) Implicit Respect: No explicitly negative statements can be identified, but neither are there explicit positive statements. (2) *Respect (balanced)*: Both, positive and negative respect is equally expressed. *(3) Explicit Respect*: This code is assigned if there is at least one explicitly positive statement about the groups and either are negative statements completely absent or positive statements are clearly dominating the negative statements.	(0) max < 2 (1) max ≥ 2
Questioning	(0) No (or only rhetorical) question posed. (1) Question asked (informational or critical).	(0) max = 0 (1) max = 1
Story-telling	(0) No story-telling. (1) Story-telling.	(0) max = 0 (1) max = 1

Note: In a first step, each speech act was coded according to the coding scheme provided in the second column of the table (see Table A.2 for coding examples). Afterwards, we aggregated the speech level data to the individual level by extracting individuals' maximum performance on each indicator. In a last step, we set the threshold values as detailed in the third column of the table.

Table A.2: DQI Coding Examples

Justification rationality (level of justification) This indicator captures whether and to what extent speakers provide reasons to back their claims.	
(0) *No Justification:* The speaker does not present any argument or only says that X should or should not be done, but no reason is given.	a) „Ich hätte eine Frage an die Deutschen. Sind doch die Einwanderer der dritten Generation bereits in Deutschland. Und werden diese dann immer noch als Einwanderer betrachtet, obwohl sie die deutsche Staatsangehörigkeit haben und in Deutschland geboren sind?“ *(spanische Frau stellt eine Frage, Gruppe 17) [Deutsche Übersetzung des Originalbeitrages durch Europolis Übersetzer]* a) „I would like to pose a question to the Germans: Now you have the third generation of immigrants in Germany. Are they still regarded as immigrants despite the fact that they possess the German citizenship and are born in Germany?” *(Spanish women poses a question, group 17) [own English translation from German transcript; German translation of the original speech was provided by Europolis translators]* b) « Moi je suis pour ça enfin [contrôles rigoureux]. Pour moi c’est bon. » *(homme français sans justifier sa position, groupe 4)* b) “This is what I favor [strict border controls]. To me, this sounds like a good solution.” *(French man does not justify his position, group 4) [own English translation from French transcript]* c) “I think that in Poland, there are not so many immigrants. I guess most of them are coming to my country from Ukraine and Belarus. However, people are also transiting Poland to get to other EU countries that have a better standard of living.” *(Polish men describing the immigration situation in his home country, group 16) [English translation from Polish transcript]*
(1) *Inferior Justification:* Here a reason Y is given why X should or should not be done, but no linkage is made between X and Y	« (…) 15 millions d’immigrés sont venus en Allemagne de manière légale - sans qualification 44 pourcent, sans qualification professionnelle, entre 22 et 40 ans : 44 pourcents, et sans qualification professionnelle en provenance de Turquie : 72 pourcent. Alors, lorsqu’on compare par exemple l’immigration en Australie. Si je veux immigrer en Australie, je dois avoir beaucoup d’argent, je dois prouver que j’ai un employeur en Australie.. Donc il faudrait

	peut-être voir ce qui se passe en dehors de l'Europe, comment on gère l'immigration dehors de l'Europe. » *(homme allemand, groupe 21) [traduction française de la contribution originale faite par les traducteurs d'Europolis]* "(…) 15 millions of immigrants came to Germany in a legal way. Without qualification 44 percent, without professional qualification, between 22 and 40 years old: 44 percent. And from Turkey, 72 percent arrived without professional qualification. Let's compare this to immigration in Australia: if I want to migrate to Australia, I need to have a lot of money, I need to prove that I do have an employer in Australia.. Therefore, we should maybe turn to countries from outside Europe and see how they handle immigration." *(German men, group 21) [own English translation from French transcript, French translation of the original speech was provided by Europolis translators]*
(2) *Qualified Justification:* A linkage is made why one should expect that X contributes to or detracts from Y.	« Le fait qu'ils quittent leur pays c'est déjà triste peut-être. Mais si dans leur propre pays, ils pourraient déjà bénéficier de meilleurs conditions de vie, cela irait bien mieux que devoir, comme des réfugiés devoir partir, quitter toute la famille, aller au très grand voyage et partir dans un pays tout à fait étranger. Donc je trouve que les pays de l'Union Européenne devraient, entre eux, centrer d'aide pour essayer de maintenir les gens dans leur propre pays en améliorant justement les conditions de vie dans leur propre pays. Alors donc, cela pourrait éviter le fait de devoir obligatoirement encore émigrer. (…) » *(femme allemande, groupe 21) [traduction française de la contribution originale faite par les traducteurs d'Europolis]* "The fact that they leave their home country is already sad, perhaps. If they could already profit from better living conditions in their own country, this would definitely be better than having to leave the family, undertake a very long journey to a country that is very different from one's own, as the refugees currently do. This is why I think that the countries of the European Union should, among each other, center the aid in order to try to keep the people in their own country by improving the living conditions in the respectable countries. This could prevent them from being forced to emigrate. (…)" *(German women, group 21) [own English translation from French transcript, French translation of the original speech was provided by Europolis translators]*
(3) *Sophisticated Justification:* At least two complete justifications are given, either two complete	« Donc d'abord sur le sujet de la confession. C'est ce que je voulais dire tout à l'heure. En disant, est-ce qu'on doit s'adapter parfaitement au pays dont qu'on arrive ou est-ce qu'au contraire, on peut avoir un équilibrage? En France, c'est très différent qu'avec l'Espagne, c'est très différent qu'avec la plupart des autres pays européens.

justifications for the same demand or complete justifications for two different demands.	Nous avons une exigence de la laïcité, c'est-à-dire que nous n'avons pas… il y a une pratique religieuse, chrétienne, historique, qui est toujours active mais en place de la laïcité au dessus dans le sens où le culte religieux est de l'endroit privé. Donc, on ne peut pas, par exemple tenir ce discours, de dire que les immigrés qui viennent en France doivent respecter une tradition chrétienne, par exemple. C'est quelque chose qu'on – on peut le dire, mais bon, sauf que dans l'effet, ça ne se présente pas dans les écoles, sauf dans les écoles privés. Mais on n'y est pas confronté. Donc pour moi, la réponse au problème, à l'éventuel problème posé par des immigrés de confession musulmane, elle passe par là en faite, la laïcité c'est ce qui permet de respecter les différences de chacun. Bon, ça c'est peut-être aussi de mes conditions mais quand vous parlez de s'adapter, moi personnellement, je m'adapte assez peux aux traditions français, donc et pourtant, je suis français, je ne vais pas partir de moi. Voilà. Donc cette histoire de se dire qu'on doit se confronter à une certaine culture – (…) Moi personnellement, je ne pense pas que ça serait la bonne réponse, parce qu'il y a trop des différences et je ne pense pas qu'un pays égale une culture en fait. Voilà, personnellement.» *(homme français, groupe 10)* "First, on the subject of religious denomination. That is what I just wanted to say before. On the question whether one needs to completely adapt to the host country or whether, on the contrary, there could be a balance. In France, the situation is very different from Spain, and different from the majority of European countries. We have the principle of laicism, this means we do not have.. there is religious practice, Christian, historically, which is still active but it is situated below the principle of laicism in the sense that religion is a private affair. This means that we cannot conduct a discourse that demands, for example, that the immigrants coming to France have to respect the Christian tradition. This is something that – one can say that but this does not apply to schools, except to private schools. But we are not confronted with that. For me, the answer to the problem, the eventual problem that Muslim immigrants pose, goes through laicism and this is what allows respecting individual differences. Alright then – this perhaps only applies to me but personally, when you speak of the need for adaptation, I don't adapt myself to the French culture in a sufficient way and I am French nevertheless, and I won't deviate from myself. The story that one needs to conform to a certain culture – (…) Personally, I don't think that this is the right answer because there are too many differences and I don't think that a country equals a culture. That's the way I see it. (…)" *(French man, group 10) [own English translation from French transcript]*

Common good orientation (content of justification) In the eyes of many deliberative democrats, good reasons should entail a common good orientation (e.g., Rawls 1971). The original DQI measures whether arguments are cast in terms of narrow group or constituency interests, whether there is neutral reference or mixed reference (i.e., reference to both narrow group interest and common good), or whether there is a reference to the common good.	
(0) Explicit statement concerning constituency or group interests is made *(here: own country)*	„(...) Meiner Meinung nach würde ich es sehr begrüssen, wenn wir vermeiden könnten, dass sich das [Einwanderung und damit verbundene Probleme] auch nach Osteuropa ausweitet. (...)“ *(litauischer Mann, Gruppe 20) [Deutsche Übersetzung des Originalbeitrages durch Europolis Übersetzer]* “(...) Personally, I would very much appreciate if we could prevent that this [immigration and connected concerns] also extends to Eastern Europe. (…)” *(Lithuanian men, group 20) [own English translation from German transcript; the German translation of the original speech was provided by Europolis translators]*
(1.5) Explicit statement in terms of both, Europe and own country is made	“To tell you the truth: in Holland, immigration is perceived as something that lingers in the minds of people and that is present in any context of discussion. The EU should take care of it, so that other countries, which do not have that kind of problems yet, such as Poland for instance, would contribute in order that we can solve this problem together. So that at least we could share the knowledge and experience in this matter. I think we should cooperate like that.” *(Dutch man, group 23) [own English translation from Polish transcript; the Polish translation of the original speech was provided by Europolis translators]*
(2) Explicit statement in terms of a conception of the common good in utilitarian or collective terms is made *(here: EU; Europe)*	« (…) Nous avons besoin d’immigrés pour le problème de démographie. Nous devons de plus en plus vieux, en Europe nous avons besoin de beaucoup des personnes. (…) Et bon je reste d’avis dès lors qu’en tant que l’Europe, et c'est pourquoi il faut qu’il est vraiment une politique européenne et non pas par pays comme ça se fait maintenant, parce que sinon on l’en sortira jamais. (…) » *(homme hollandais, groupe 3) [traduction française de la contribution originale faite par les traducteurs d’Europolis]* “(…). We need immigrants in order to solve the demographic problem. We grow older and older in Europe and we need a lot of people. (…) Well, I hold on the position that from a prospective view of Europe, this is why we need a European

	policy instead of state based solutions like we have them now. If we don't opt for a European solution, we will never overcome this problem. (…)" *(Dutch men, group 3) [own English translation from French transcript; the French translation of the original speech was provided by Europolis translators]*
(3) Explicit statement in terms of the difference principle is made *(solidarity, quality of life, global justice, etc.)*	„Ja ich bin also der Meinung, dass die Aufgabe der EU das eigentlich nicht unbedingt ist, die Länder und die Grenzen dicht zu machen, weil wir müssen das alles ja doch auch ein bisschen globaler sehen, wir sind doch nun einmal eine Welt und die Welt besteht doch nun einmal aus vielen anderen Ländern und wir müssen doch dann auch dafür sorgen, dass es der restlichen Bevölkerung und den anderen Menschen in irgend einer Form gut geht, die wir über all die Jahre ausgebeutet haben. Wir haben doch eine Verpflichtung; also Menschlichkeit ist doch ein ganz oberes Prinzip und ich bin einfach der Meinung, wir können uns nicht abschotten und können einfach unser Süppchen alleine kochen. (...)" *(deutsche Frau, Gruppe 9)* "In my opinion, it should not be the task of the EU to close the borders, because we need to consider this from a more global point of view. It's a fact that we are just one world and the world consists of many other countries and we have to see that the rest of the population and the other people that we exploited for so many years are doing well in any way. We do have a responsibility. Humanity is the highest principle and I think that we cannot seal ourselves off and 'cook our soup' unilaterally. (…)" *(German women, group 9) [own English translation from German transcript]*
Respect and interactivity The updated version of the DQI allows capturing interactivity, a hitherto neglected aspect in the empirical measurement of deliberation. The introduction of a neutral response category ensures that every uptake of others' arguments is encompassed. This includes uptakes in the form of questioning, recapitulations of what others have said, and references to stories. Note that expressing disagreement is not bound to the code of negative respect. As examples (2a) and (3c) demonstrate, disagreement can also be expressed in a neutral or even positive way. However, when people explicitly mentioned that they are in full disagreement with other participants' arguments (without acknowledging the other's position in any respect), this was coded as a disrespectful statement (0a).	
(0) Negative (disrespectful) reference to other participants'	a) « Et que les deux premiers avis, premiers opinions soient favorable à ce que cette politique soit défini par les états et non par l'Union Européenne. Alors je suis tout à fait en désaccord. Je pense que la politique de migration doit être

arguments	définie dans la base par les organes Européens. (…) » *(homme portugais répond aux personnes précédentes, groupe 1) [traduction française de la contribution originale faite par les traducteurs d'Europolis]* a) "And the first two opinions that were expressed were in favor of regulating this policy at the country level instead of the level of the European Union. I am completely opposed to that. I think that in its main features, the European institutions should define the immigration policy. (…)" *(Portuguese man responds to previous speakers, group 1) [own English translation from French transcript; the French translation of the original speech was provided by Europolis translators]* b) « Je ne vois pas au quoi l'immigration illégale pose un problème de logement pour les français ou des choses comme ça. Je crois qu'il ne faut pas faire attention, pas tous mélanger. (…) Je crois que ce n'est pas possible de faire une confusion de cette sorte là. » *(homme français répond à la personne précédente, groupe 21)* b) "I don't see in which way illegal immigration presents a problem of housing for the French people. I think we should not pay attention, should not mix everything up. (…) I think it is not possible to confuse things like that." *(French man responds to a previous speaker, group 21) [own English translation from French transcript]*
(2) Neutral reference to other participants' arguments	a) „Also an der Stelle möchte ich meinen beiden Vorrednern in gewisser Weise widersprechen. (...)" *(deutscher Mann, Gruppe 24)* a) "At this point, I would like to contradict the previous speakers to some extent. (…)" *(German man, group 24) [own English translation from German transcript]* b) "According to what the previous speaker said, immigrants are coming because there are jobs that Polish people do not want to take, so in my country, the immigrants are coming to take these jobs over. (…)" *(Polish women, group 16) [own English translation from Polish transcript]*
(3) Positive (explicitly respectful) reference to other participants' arguments	a) « Je suis tout à fait d'accord avec [l'homme français]. S'il y a des lois, il faut les respecter. (…) » *(homme allemand, groupe 4) [traduction française de la contribution originale par les traducteurs d'Europolis]* a) "I fully agree with [French man]. If there are laws, they need to be respected. (…)" *(German man, group 4) [own English translation from French transcript; the French translation of the original speech was provided by Europolis*

	translators] b) « Je pense que la proposition de [l'homme allemand] concernant le traitement différenciel des personnes est très importante. (…) » *(homme allemand, groupe 4) [traduction française de la contribution originale faite par les traducteurs d'Europolis]* b) "I think that the proposal of [German man] concerning a differentiated treatment of people is very important. (…)" *(German man, group 4) [own English translation from French transcript; the French translation of the original speech was provided by Europolis translators]* c) „Zu deinem Kommentar noch; <u>das stimmt, aber</u> das erste ist auch ausschlaggebend; es ist ja auch ein gewisser Prozentsatz da, die wollen sich nicht integrieren. (...)" *(deutscher Mann, Gruppe 9)* c) "On your comment: <u>you are right but</u> the first point is crucial as well; there is also a certain amount of people that doesn't make an effort to integrate. (…)" *(German man, group 9) [own English translation from German transcript]*
Respect towards groups (migrants) Respect toward groups implies that participants, either implicitly or explicitly, acknowledge the needs and rights of different social groups. In the context of Europolis discussions on third country migration, migrants from outside the EU are the social group of interest.	
(0) *No Respect*: This code is reserved for speeches in which there are only or predominantly negative statements about the groups.	a) « L'immigration du 20ème et 21ème siècle, c'est l'immigration des gens des cultures complètement opposées à la nôtre. Et c'est ça le problème. Ces gens-là, ils ne s'attachent pas du tout à notre culture. (…) On a eu des problèmes, ces gens-là sont différents. » *(homme français, groupe 4)* a) "The immigration of the 20th and 21st century is an immigration of people from cultures that are completely opposed to our. And this is the problem. These people do not adapt at all to our culture. (…) We have had problems – these people are different." *(French man, group 4) [own English translation from French transcript]* b) „(...) Also Leute die nach Europa kommen, wollen zumeist definitiv hier bleiben, nicht irgendwie nur ein Zeitvertrag ableisten und dann wieder weggehen. Sie wollen ins europäische System eintreten. Sie wollen nicht wirklich sich einfügen in die Kultur, aber die wirtschaftlichen Vorteile nutzen und nicht wirklich als Europäer auftreten dann. Das ist

	sehr schwer, sie dann wieder wegzuschicken, man findet sie kaum. Ja ich weiss, dass es schwierig ist da. Aber ich weiss auch, dass es viele Einwanderer gibt, von denen kommen sogar die meisten als Touristen. Mit einem Monatsvisum und dann bleiben sie hier als illegale Einwanderer. (...)“ *(spanischer Mann, Gruppe 19) [Deutsche Übersetzung des Originalbeitrages durch Europolis Übersetzer]* b) „People that come to Europe intend, most of the times, to stay here on a long-term base. They don't want a temporal contract that requires them to leave after the contract expires. They want to enter the European system. They don't make serious efforts to adapt to our culture, but they want to profit from the economic benefits without becoming European. It is difficult to make them leave again because you hardly find them. Yes I know that they have it difficult over there. But I also know that many of the immigrants arrive as tourists with a monthly visa. After the expiration of the visa they just stay and become illegal immigrants. (…)” *(Spanish man, group 19) [own English translation from German transcript; the German translation of the original speech was provided by Europolis translators]*
(2) *Respect (balanced)*: Both, positive and negative respect is equally expressed.	„(...) Man kann das nicht alles über den Kamm scheren und alle nach einem Schema beurteilen, denn es gibt Einwanderer, die in andere Länder eindringen um vielleicht sich der Arbeit zu entziehen oder sich unrechtmässig zu bereichern durch Diebstahl und so weiter; es sind also Parasiten. Aber es gibt daneben auch Einwanderer, die sich anstrengen um sich den Gegebenheiten und insbesondere auch der Kultur des Landes anzupassen.“ *(polnische Frau, Gruppe 9) [Deutsche Übersetzung des Originalbeitrages durch Europolis Übersetzer]* “(…) There is no way in trying to apply one standard to all. There are immigrants who infiltrate a country, maybe in order to avoid having to work or in order to enrich themselves illegally via theft and so on – these are parasites. On the other hand, there are immigrants that make an effort in order to adapt themselves to the circumstances of a country and in particular to its culture.” *(Polish woman, group 9) [own English translation from German transcript; the German translation of the original speech was provided by Europolis translators]*
(3) *Explicit Respect*: This code is assigned if there is at least one explicitly positive statement about the groups and either are negative statements completely absent or positive statements are	a) « (…) Je pense même que parmi les autochtones, il y a plus des gens qui ne veulent pas travailler que parmi les immigrants. (…) » *(homme hollandais, groupe 3) [traduction française de la contribution originale faite par les traducteurs d'Europolis]* a) “(…) I think that among the indigenous population, there are more people who don't want to work than there are among immigrants. (…)” *(Dutch man, group 3) [own English translation from French transcript; the French translation*

clearly dominating the negative statements.	*of the original speech was provided by Europolis translators]* b) „(...) Ich denke, niemand verlässt sein Heimatland aus Spass, sondern die Leute machen das aus Not, weil sie nichts mehr haben um zu überleben. Kaum noch etwas zu essen. Sie machen das nicht unbedingt freiwillig. Also das Leben muss schon sehr schlecht sein, es muss einem sehr schlecht gehen, damit man sich dazu entschliesst, seine Heimat zu verlassen und dass es eben auch keinerlei Aussicht auf Besserung gibt, dass sich die Dinge eben nicht kurzfristig im Land verbessern lassen. (...)“ *(spanischer Mann, Gruppe 20) [Deutsche Übersetzung des Originalbeitrages durch Europolis Übersetzer]* b) “(…) I think no one leaves his home country just for fun. Rather, people leave because of misery, because they don’t have anything left in order to survive. Hardly anything to eat. They don’t necessarily do that on a voluntary basis. Life must treat one very bad in order to make the decision to leave one’s home country. There must be no prospect of improvement, there must be a certainty that these things will not change for the better in the near future.” *(Spanish man, group 20) [own English translation from German transcript; the German translation of the original speech was provided by Europolis translators]*
Questioning The updated version of the DQI regards inquisitiveness as an essential aspect of an interactive deliberative process. In the context of Europolis, we do not distinguish between informational and critical questions (see Ikuenobe 2001).	
(1) Question asked (informational or critical)	a) „Eine Frage an [deutschen Mann]: In welche Dimension wächst das Problem mit der Abwanderung von Arbeitskräften?“ *(deutsche Frau, Gruppe 9)* a) “I have a question to [German man]: To what extent increases the problem of emigration for working purposes?” *(German women, group 9) [own English translation from German transcript]* b_1) „Ich möchte mal die Dame nebenan fragen, was sie macht, wenn sie ungebetenen Besuch kriegt; wenn sie heimkommt und in ihrer Wohnung ist jemand drin. Was sie dann macht.“ *(deutscher Mann, Gruppe 9)* b_2) „Da würde ich fragen, wie sie die Wohnung, also wirklich intimen Lebensraum, mit dem breiten Raum eines Landes oder einer weiteren Umgebung vergleichen können?“ *(deutscher Mann, Gruppe 9)*

	b_1) “I’d like to ask the lady next to me what she would do when she would receive unsolicited visit. When she returns and there is somebody, right inside her apartment. What would she do?” *(German man, group 9) [own English translation from German transcript]* b_2) “In this case I’d like to ask you how you can compare an apartment, meaning intimate living space, with a country where there is far more space.” *(German man, group 9) [own English translation from German transcript]*
Story-telling According to Polletta und Lee (2006), “story-telling” is the most important component of alternative forms of communication. Therefore, the updated version of the DQI captures the presence of stories. In Europolis, story-telling ranges from the description of a situation in one’s home country to the report of a personal narrative.	
(1) Story-telling	a) „(...) In C, da gibt es zum Beispiel ein Ort, da sind in den gleichen Schulen muslimische Kinder und da mussten wir in den Schulen zum Beispiel das ganze Menü in der Kantine ändern, weil sie zum Beispiel kein Schweinefleisch essen. (...)“ *(spanische Frau, Gruppe 20) [Deutsche Übersetzung des Originalbeitrages durch Europolis Übersetzer]* a) “(...) In C for example, there is a place where Muslim children go to the same schools as Spanish children. There, we were forced to change the whole menu of the cafeteria because they don’t eat pork, for example. (…)” *(Spanish woman, group 20) [own English translation from German transcript; the German translation of the original speech was provided by Europolis translators]* b) « Je voudrais rapidement réagir, donc j’ai une bonne expérience, c'est-à-dire donc une famille, en provenance de Mongolie, ils sont revenu et donc il y a maintenant une des filles qui a fait un doctorat en langue allemande et donc un autre fils a fait un examen en allemand, donc je ne connais aucune personne où on pourrait dire voilà, il a juste appris par cœur et puis il a oublié tout, pour avoir ce certificat [de nationalité]. Donc j’ai juste eu la chance d’avoir que des bonnes expériences. » *(homme allemand, groupe 4) [traduction française de la contribution originale faite par les traducteurs d’Europolis]* b) “I’d like to quickly reply by providing a positive experience. There is a family, originating from Mongolia, who returned and one of the daughters now has a PhD in German language and one of their sons did an exam in German. This is to say that I don’t know any person from whom I could say that ‘here we go – he just learned it by heart and after

	receiving the [citizenship] certificate, he forgot everything again'. In sum, I only had the chance to make positive experiences." *(German man in reply to someone claiming that immigrants attain their citizenship certificate by learning everything required by heart, group 4) [own English translation from French transcript; the French translation of the original speech was provided by Europolis translators]*

Note: The audio discussions of the coded groups were available either in French, German, or Polish. For the French and German groups, the coding was based on the transcripts from the audio discussions. Thus, for non-German or non-French speakers (or German speakers in French groups as well as French speakers in German groups), our coding is based on the translation of the original speech act as provided by Europolis translators. For the two Polish groups, our coding relies on an English translation of the respective group discussions. In the present Table, our coding is always based on the first example provided in the Table. The English translations of the German and French speeches only serve purposes of coding demonstration.

Appendix B

INTER-CODER RELIABILITY

A test of inter-coder reliability was conducted prior to coding the full database. Three coders independently evaluated 80 speeches. Four statistics were calculated: the ratio of coding agreement (RCA; Holsti, 1969) which simply indicates the percentage of observations that were coded in the same way by coders for a given indicator; Cohen's (1960) κ ("kappa") and, where appropriate, Spearman's rank correlation (Siegel, 1956) and the standardized α (a reliability statistic; see Holsti, 1969). These statistics take into account the difference between actual coding agreement and the level of coding agreement that could be expected to occur by chance (see also Steenbergen et al., 2003).

We consider both the overall performance of the DQI and its performance for each of the coding categories. The three coders agreed on average in 85.4% of a total of 480 rendered judgments.[49] The corresponding value for Cohen's Kappa κ mounts to 0.70. Where appropriate, Spearman's *r* and the standardized α were calculated (r=0.81; α=0.89). All findings indicate substantial agreement.

Agreement on Specific Categories: Table B1 reports the reliability statistics for specific coding categories. First, consider the justification rationality (level of justification), *RCA*=0.65, which looks respectable at first glance. Taking into consideration that the coders may have agreed by chance alone, we also computed Cohen's κ, which equals 0.52. This is indicative of moderate agreement and clearly misses the benchmark for substantial agreement (Landis and Koch, 1977). Since justification rationality is an ordinal indicator, it is also useful to consider Spearman's rank correlation, which accounts for the difference in the rank orderings of speeches between coders. Spearman's *r* equals 0.78, which produce a standardized item α of 0.92. These results indicate that the low coder agreement reported for RCA and Kappa rather resulted from minor discrepancies about which code on the scale should be applied.

[49] For each coding category (justification rationality, pursuit of self-interest, EU's interests and world interests, respect toward participants' arguments, and story-telling), 80 judgments were rendered.

Table B.1: Inter-Coder Agreement per Category

	RCA	*Kappa*	*s.e.*	*Spearman*	*Alpha*
Justification rationality (Level of justification)	0.654	0.516**	0.068	0.778**	0.917
Country interest (Content of justification 0)	0.950	0.721**	0.110		
European Union interests (Content of justification 2)	0.925	0.642**	0.111		
World community interests (Content of justification 3)	0.975	0.592**	0.096		
Respect toward participants' arguments	0.792	0.698**	0.073	0.813**	0.913
Respect toward groups (migrants) [a]	0.850	0.723**	0.113	0.756**	0.833
Story-telling	0.946	0.852**	0.086		

Note: N=80. **p<0.01. [a] This category has been coded at a later point in time. Therefore, a separate reliability test was conducted where two independent coders coded a total number of 40 speeches.

For the content of justification, we ran three different tests since we need to distinguish among references to country interest (0), the European Union (2), and the world community (3). For the first content of justification, the coders agreed in 95% of the time. Taking chance agreement into account, κ=0.72; this is statistically significant and indicates substantial agreement. For the second content of justification coding, the coders agreed in 92.5% of the cases. However, κ is at 0.64 and only slightly above the benchmark of 0.61 for substantial agreement (Landis and Koch, 1977). For the third content of justification coding, the coders agreed in 97.5% of the cases. However, the fact that that κ = 0.59 indicates that coders only agreed that references to the world community occur rarely, but then disagreed on what finally deserves a positive code. While the scores for the average score hover around the benchmark for substantial agreement, they are more daunting if we consider the wide range of inter-coder agreement among the pairs of coders.

Therefore, we performed another round of coding. After the disagreement for level and content had been settled and disputed cases were recoded, a second reliability test was performed for level and content of justification (Table B.2). The re-test indicates substantial coder agreement after re-coding both for level and content of justification.

The reliability of the respect indicator, in turn, was satisfactory. Regarding respect toward other participants' arguments, coders agreed on 79.2% of the codes. Furthermore, κ=0.70,

which is again significant and indicative of substantial agreement. Since the indicator for respect toward participant's arguments is ordinal, Spearman's rank correlation was also computed: $r = 0.81$, which results in an impressive α of 0.91. Similarly positive results are obtained for the indicator of respect towards groups (migrants). Since this category has been coded at a later point in time, a separate reliability test was conducted (see note in Table B.1).

The final coding category is story-telling. Here, *RCA* = 0.95, which is indicative of excellent agreement. κ equals .85, which is significant and reflects high levels of agreement.[50]

In sum, these results show that reliable measurement through the DQI in citizen deliberation is possible. Even the worst reliability scores are still respectable, suggesting that different coders looking at the same transcripts will be able to agree on the DQI and its components. This is an important result, as it greatly increases the confidence one can place in the analyses performed with the DQI indicators.

Table B.2: Inter-Coder Agreement Re-Test (Justification)

	RCA	*Kappa*	*s.e.*	*spearman*	*alpha*
Justification rationality (Level of justification)	0.800	0.725**	0.128	0.942**	0.974
Country interest (Content of justification 0)	0.967	0.917**	0.222		
European Union interests (Content of justification 2)	1.000				
World community interests (Content of justification 3)	1.000				

Note: N=20. **p<0.01.

[50] Since we did not distinguish between informational and critical questions, no reliability test was conducted for the category of questioning.

Appendix C

SUMMARY STATISTICS OF THE DQI CODING

Table C.1: Summary Statistics of the Overall Discourse Quality

	Level of speeches [b]	*Level of participants* [c]
Justification rationality		
No justification	18.5%	2.8%
Inferior justification	39.2%	18.9%
Qualified justification	32.2%	41.7%
Sophisticated justification	10.0%	36.6%
Common good orientation		
No explicit notion of interests	75.5%	30.6%
Country interests	8.8%	20.6%
Mixed (country/EU)	4.6%	12.2%
European Union interests	7.3%	22.2%
World community interests	3.8%	14.4%
Respect towards other participants arguments [a]		
Negative reference to arguments	5.2%	0.0%
No explicit reference to arguments	59.1%	16.7%
Neutral reference to arguments	38.2%	28.3%
Positive reference to arguments	18.6%	55.0%
Respect towards groups		
Negative respect	8.6%	2.8%
Implicit respect	73.0%	42.2%
Balanced respect	3.9%	12.2%
Positive respect	14.5%	42.8%
Story-telling	32.9%	70.6%
Questioning	12.7%	40.0%

Note: [a] Up to three references were coded per speech act. Therefore, percentages do not sum up to 100%. [b] Entries refer to the number of statements that received the respective code, compared to the total number of statements provided (N=943). [c] Entries indicate the percentage of active speakers (N=180) that reached this standard. For example, 36.6% provided sophisticated justification at least once, 41.7% never reached the level of sophisticated justification but provided qualified justification at least once, and 2.8% never provided a justification for any of their speeches.

Appendix D

PROPERTIES OF THE SELECTED SAMPLE OF DISCUSSION GROUPS

Table D.1: Group Composition of Selected Sample

Group ID	*Nationalities*	*heterogeneity*[a]	*f/g/p*[b]	*N*
1	3 French, 6 Portuguese, 6 Romanians	1.92	f	15
2	8 French, 6 Hungarians	0.98	f	14
3	2 Belgians (Dutch speaking), 7 Dutch, 6 French	0.62	f	15
4	1 Austrian, 6 French, 7 Germans	0.56	f	14
9	1 Austrian, 6 Germans, 8 Polish	1.10	g	15
10	1 Belgian (French speaking), 6 French, 7 Spanish	1.12	f	14
16	9 Polish, 6 Slovenes	0.48	p	15
17	1 Austrian, 6 Germans, 8 Spanish	1.10	g	15
19	5 Danes, 4 Germans, 7 Spanish	1.30	g	16
20	1 Austrian, 5 Germans, 5 Lithuanians, 5 Spanish	2.11	g	16
21	1 Austrian, 7 French, 6 Germans, 1 Luxembourger	0.61	f	15
23	2 Belgians (Dutch speaking), 5 Dutch, 7 Polish	1.20	p	14
24	1 Austrian, 1 Cypriot, 8 Germans, 3 Greeks	1.11	g	13

Note: [a] Index of national heterogeneity of small groups. See Table D.2 for details on the creation of the index. [b] Indicates the language of the recordings of the audio discussions from which the discussions have been coded (see also Table A.2 and corresponding note). f=French; g=German; p=Polish.

Table D.2: Comparison of Selected Sample to the Rest

	Selected Sample (N=191)		**Rest (N=157)**		
	Mean (Std.)	*N*	*Mean (Std.)*	*N*	*t-test*
Women 1=female; 0=male	0.46 (0.50)	191	0.49 (0.50)	157	0.55
Age Year of birth	1959.4 (16.67)	191	1961.9 (16.63)	156	1.41
Education Age in year when main education ended	20.88 (4.59)	171	19.91 (4.08)	133	-1.91^{+}
Working-class 1=yes; 0=no	0.22 (0.41)	187	0.26 (0.44)	155	0.96
Ideology In political matters people talk of "the left" and "the right". What is your position? Please indicate your views using any number on a scale where '0' means "left" and '10' means "right."	4.91 (2.72)	184	5.32 (2.52)	146	1.42
Salience T2 On a scale from 0 to 10, where '0' is "no problem at all", '10' is "the most serious problem we face", and '5' is "exactly in the middle", how serious a problem or not would you say immigration is?	5.72 (2.40)	187	6.61 (2.47)	153	3.39***
Knowledge T2 Amount of correct answers to the three knowledge questions on immigration available in the questionnaire. The questions asked for the definition of Blue Card workers, the role of the EU in the current immigration policy and for a figure of the number of immigrants currently living in	1.07 (0.89)	191	1.02 (0.92)	157	-0.56

the EU. For each of the questions, the participants could choose out of four possible options and an additional "don't know" category.					
Read briefing material T2 How much of the briefing material you were sent would you say you have read? (1) have not read or have just glanced at it (2) Have read less than half of it (3) Have read about half of it (4) Have read more than half of it (5) Have read all or nearly all of it	3.92 (1.21)	186	3.71 (1.31)	154	-1.53
Interest in politics T2 On a scale from 0 to 10, where '0' is "not at all", '10' is "passionately", and '5' is "exactly in the middle", how interested or not would you say you generally are in politics?	6.48 (2.22)	188	6.19 (2.53)	155	-1.11
Attitude scale on immigration T2 This variable has been built as a scale from five central questions on immigration that proved to be substantially related in a test of inter-item correlation (α=0.76). The questions ask for participants' opinions on regularization/sending back irregular immigrants, reinforcing border controls, imposing penalties on employers who hire irregular immigrants, whether immigration increases criminality, and whether amnesty given to irregular immigrants will increase irregular immigration. The (standardized) scale has been generated automatically via the inter-item correlation command in Stata. The scale ranges from -1.27 to 2.17 with lower values indicating more conservative views on immigration.	0.06 (0.05)	191	-0.07 (0.06)	156	1.66^{+}
Dominant facilitator T3 My small group moderator sometimes tried to influence the group with her/ his ideas. The scale ranges from 0 ("completely disagree") to 10 ("completely agree").	1.06 (2.38)	187	1.05 (2.15)	152	-0.05
National heterogeneity of small group This measure is based on the Ethnolinguistic Fractionalization Index (ELF) (Fearon and Laitin, 2003). ELF= $1-\Sigma s_i^2$ where s_i is the share of group *i* out of a total of *n* groups (here: nationalities).	1.10 (0.48)	191	1.25 (0.37)	157	3.36***

The ELF scores have been weighed by a factor of 3 for groups that incorporated members from all three regions (WE, SE, and CEE), by a factor of 2 for groups that incorporated members from two out of the three regions, and by a factor of 1 when only one region was represented in the group. The group scores range from 0.48 to 2.11, with higher values indicating more heterogeneous groups. The most homogeneous and the most heterogeneous groups are part of our sample.					
Opinion on item "regularizing irregular immigrants" T2 Some people think that [COUNTRY] should send all illegal immigrants back to the countries they came from. Suppose these people are at one end of a 0-to-10 scale, at point 0. Other people think that [COUNTRY] should legalize all the illegal immigrants currently here. Suppose these people are at the other end of the scale, at point 10. People who are exactly in the middle are at point 5, and of course other people have opinions at other points between 0 and 10. Where would you place your views on this scale, or haven't you thought much about that?	5.57 (2.43)	185	5.22 (2.86)	152	-1.21
Opinion on item "regularizing irregular immigrants" T3	5.72 (2.48)	185	5.47 (2.76)	152	-0.88
Opinion on item "European involvement in immigration affairs" T2 On a scale from 0 to 10, where '0' means "entirely by the individual Member States ", '10' means "entirely at the EU level", and '5' is "exactly in the middle", at what level do you think decisions should be made in the area of immigration? [a]	4.84 (3.02)	184	5.12 (3.07)	153	0.85
Opinion on item "European involvement in immigration affairs" T3	5.06 (2.76)	184	4.68 (2.90)	153	-1.23

Note: All variables are retrieved from the Europolis questionnaire dataset. T2 indicate questionnaire responses right at the start of the three-day event in Brussels, T3 indicate questionnaire responses right at the end of the three-day event. [a] For the sake of interpretability, the poles of the original question have been exchanged in this paper. This is the reversed version of the question. T-test: $^{+}p<0.1$, $*p<0.05$, $**p<0.01$, $***p<0.001$. For a comparison between test- and control-group, visit http://ccd.stanford.edu.

Appendix E

ADDITIONAL ANALYSES (CHAPTER 3)

Table E.1: Determinants of Contribution (incl. Cross-Level Interaction)

	Speaking Time		Frequency	
	Model E1	*Model E2*	*Model E3*	*Model E4*
Intercept	5.596***	5.642***	1.550***	1.550***
	(0.244)	(0.251)	(0.165)	(0.160)
Women	-0.678**	-0.657**	-0.120	-0.114
	(0.223)	(0.225)	(0.116)	(0.117)
Education	0.019	0.026	0.012	0.015
	(0.026)	(0.026)	(0.013)	(0.013)
Working-class	-0.327	-0.289	-0.267^{+}	-0.247
	(0.279)	(0.284)	(0.149)	(0.151)
Younger participants	0.044	-0.002	-0.134	-0.154
	(0.291)	(0.292)	(0.151)	(0.153)
Older participants	0.184	0.155	-0.118	-0.142
	(0.290)	(0.291)	(0.148)	(0.150)
Southern Europeans (SE)	0.109	0.041	-0.062	-0.090
	(0.319)	(0.362)	(0.181)	(0.226)
Central Eastern Europeans (CEE)	-0.724	-0.714^{+}	-0.342^{+}	-0.276
	(0.450)	(0.379)	(0.187)	(0.190)
Interest in politics	0.190***	0.177**	0.075*	0.070*
	(0.056)	(0.056)	(0.030)	(0.030)
Knowledge	0.294*	0.315*	0.039	0.040
	(0.126)	(0.126)	(0.063)	(0.064)
Salience	-0.003	0.002	-0.016	-0.015
	(0.053)	(0.054)	(0.027)	(0.027)
Attitudes (strongly progressive)	-0.202	-0.260	0.090	0.056
	(0.288)	(0.291)	(0.146)	(0.147)
Attitudes (strongly conservative)	-0.434	-0.441	0.136	0.122
	(0.280)	(0.285)	(0.148)	(0.152)
National heterogeneity of SG	0.329	0.181	-0.163	-0.138
	(0.356)	(0.350)	(0.284)	(0.289)
Heterogeneity x women	-0.759		-0.334	
	(0.466)		(0.261)	
Heterogeneity x CEE		-1.372*		-0.555
		(0.69)		(0.432)
Heterogeneity x SE		-0.049		-0.062
		(0.728)		(0.467)
Random Effect: Individual level	1.931	1.960	0.241	0.251
	(1.390)	(1.400)	(0.491)	(0.501)
Random Effect: Group level	0.610	0.284	0.144	0.117
	(0.781)	(0.532)	(0.380)	(0.342)
AIC	654.9	654.2	364.8	366.8
Log Likelihood	-310.4	-309.1	-165.4	-165.4

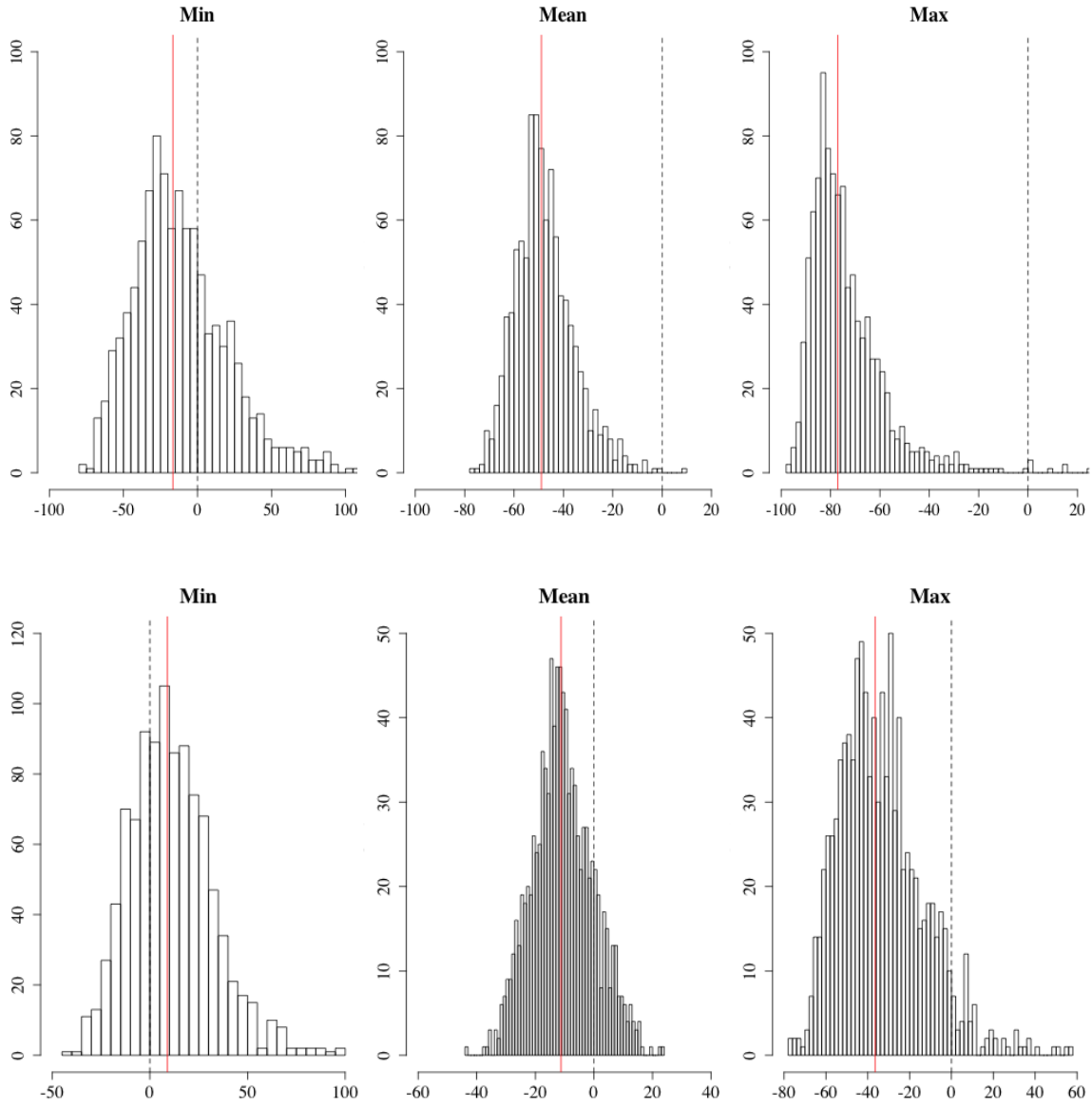
Notes: Multilevel models are estimated with the lme4 package in R (Bates et al., 2011). For contribution in speaking time, a linear hierarchical model with varying slopes for CEE participants is estimated. The dependent variable is operationalized as log of the total amount of speaking time. For frequency of contribution, a lognormal Poisson model is estimated (varying intercept) in order to account for overdispersion (Bolker, 2010; Bates et al., 2011). All ordinal and continuous predictor variables are centered at the mean. Estimated standard errors are shown in parentheses. $^{+}$p<0.1, *p<0.05, **p<0.01, ***p<.001. Individual N=173. Group N=13.

Table E.2: Determinants of Consideration (incl. Cross-Level Interaction)

	Model E5	*Model E6*
Intercept	0.940***	0.937***
	(0.153)	(0.151)
Women	-0.172	-0.135
	(0.133)	(0.134)
Education	0.011	0.016
	(0.014)	(0.015)
Working-class	-0.505**	-0.488**
	(0.179)	(0.181)
Younger participants	-0.148	-0.144
	(0.172)	(0.175)
Older participants	-0.266	-0.304^{+}
	(0.168)	(0.172)
Southern Europeans (SE)	-0.312	-0.469^{+}
	(0.201)	(0.247)
Central Eastern Europeans (CEE)	-0.186	-0.180
	(0.201)	(0.202)
Interest in politics	0.014	0.001
	(0.034)	(0.035)
Knowledge	0.011	0.010
	(0.071)	(0.072)
Salience	-0.018	-0.014
	(0.030)	(0.031)
Attitudes (strongly progressive)	0.195	0.148
	(0.162)	(0.164)
Attitudes (strongly conservative)	0.432**	0.434**
	(0.161)	(0.168)
National heterogeneity of SG	0.026	-0.137
	(0.240)	(0.243)
Heterogeneity x women	-0.685*	
	(0.320)	
Heterogeneity x CEE		-0.518
		(0.432)
Heterogeneity x SE		0.422
		(0.517)
Random effects		
Individual level	0.141	0.164
	(0.376)	(0.405)
Group level	0.056	0.033
	(0.237)	(0.182)
AIC	279.8	283.8
Log Likelihood	-122.9	-123.9

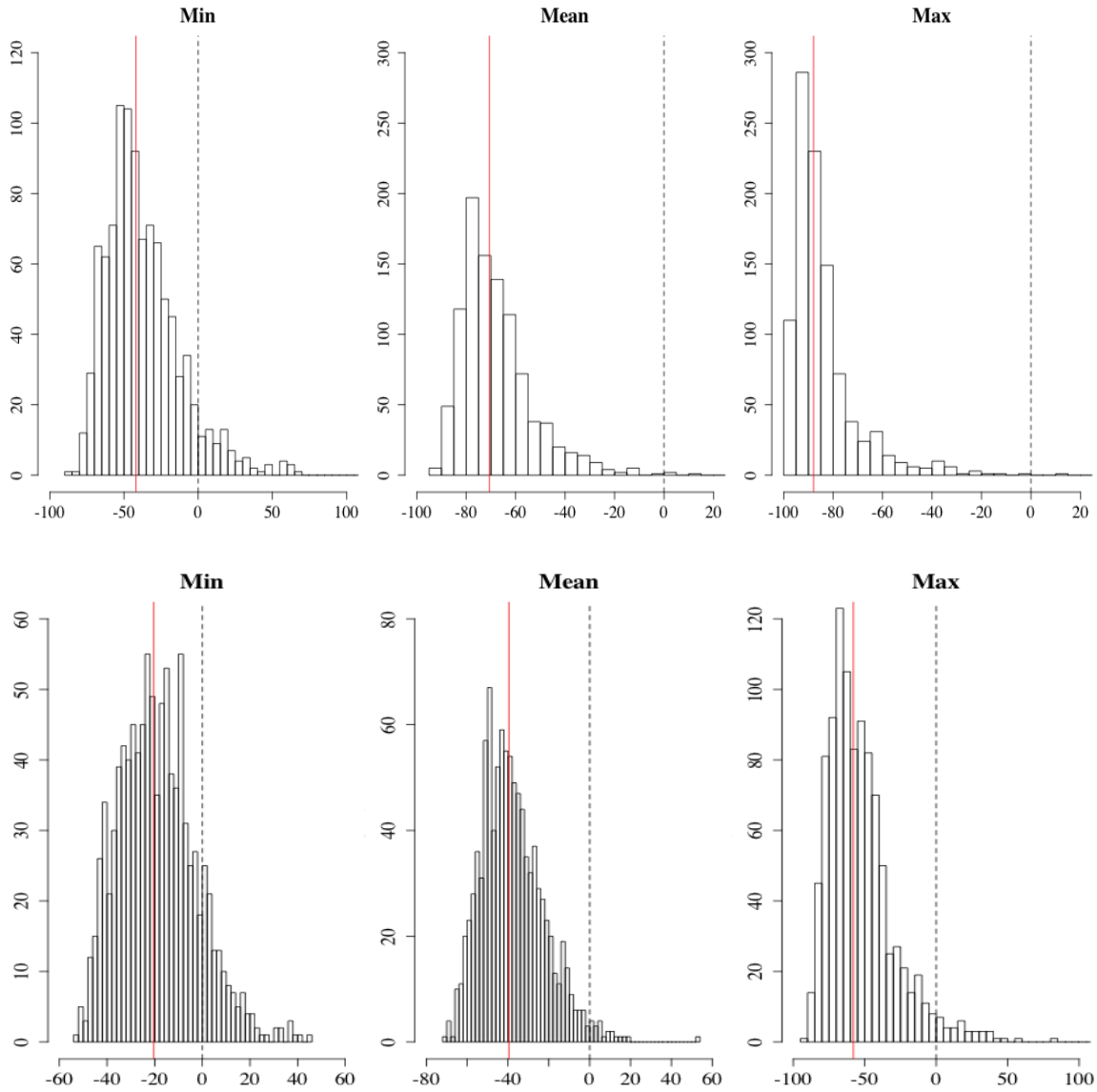
Notes: Multilevel models are estimated with the lme4 package in R (Bates et al., 2011). Lognormal Poisson models are estimated (varying intercept) in order to account for overdispersion (Bolker, 2010; Bates et al., 2011). All ordinal and continuous predictor variables are centered at the mean. Estimated standard errors are shown in parentheses. $^{+}$p<0.1, *p<0.05, **p<0.01, ***p<.001. Individual N=163. Group N=13.

Figure E.1: Simulated Effect of Women on Contribution at Different Levels of National Heterogeneity



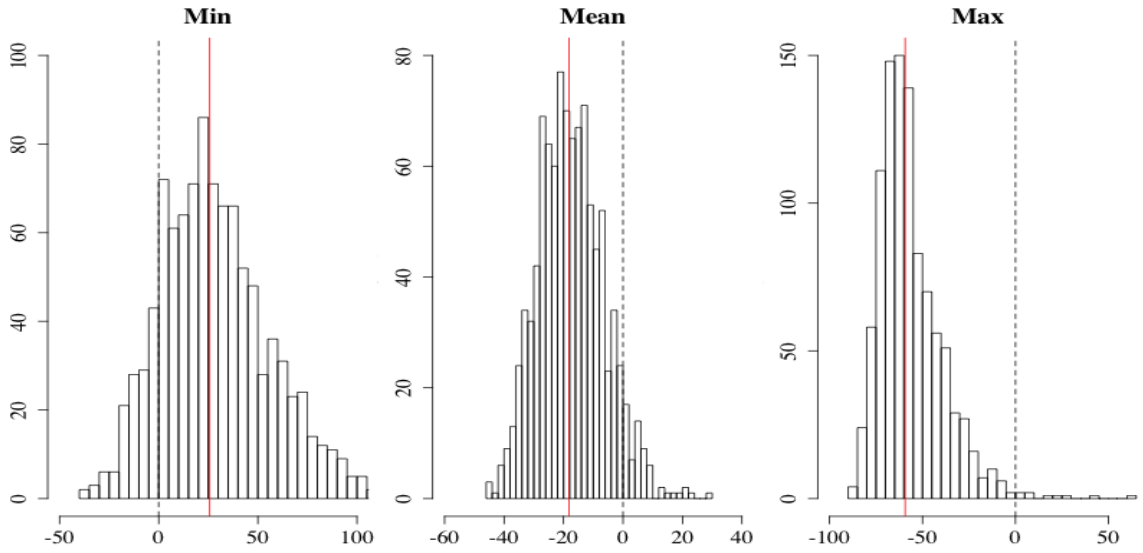
Note: Simulations are based on the statistical models presented in Table E.1. The models were simulated in R (1000 draws). The first row shows the frequency distribution of the simulations for total amount of speaking time (Model E1), and the second row shows the distribution of the simulations for frequency of contribution (Model E3). The x-axis indicates the percentage difference in contribution that is attributed to the effect of being female, at the corresponding level of national heterogeneity and by holding all other factors constant.

Figure E.2: Simulated Effect of CEE Participants on Contribution at Different Levels of National Heterogeneity



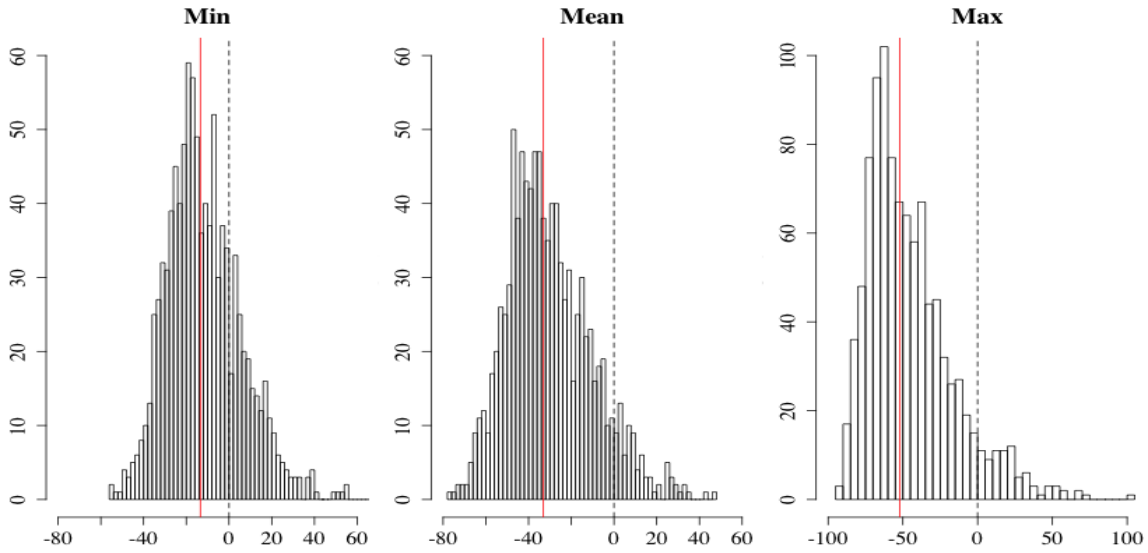
Note: Simulations are based on the statistical models presented in Table E.1. The models were simulated in R (1000 draws). The first row shows the frequency distribution of the simulations for total amount of speaking time (Model E2), and the second row shows the distribution of the simulations for frequency of contribution (Model E4). The x-axis indicates the percentage difference in contribution that is attributed to the effect of belonging to a CEE member state (compared to belonging to a WE member state), at the corresponding levels of national heterogeneity and by holding all other factors constant. Note that the minimum, mean and maximum value of national heterogeneity have been recalculated. Calculations solely base on the groups in which CEE participants were engaging with participants from WE and/or SE member states.

Figure E.3: Simulated Effect of Women on Consideration at Different Levels of National Heterogeneity



Note: Simulations are based on Model E5 presented in Table E.2. The model was simulated in R (1000 draws). The x-axis indicates the percentage difference in consideration that is attributed to the effect of being female, at the corresponding level of national heterogeneity and by holding all other factors constant.

Figure E.4: Simulated Effect of CEE Participants on Consideration at Different Levels of National Heterogeneity



Note: Simulations are based on Model E6 presented in Table E.2. The model was simulated in R (1000 draws). The x-axis indicates the percentage difference in consideration that is attributed to the effect of belonging to a CEE member state (compared to belonging to a WE member state), at the corresponding levels of national heterogeneity of the group discussions and by holding all other factors constant. Note that the minimum, mean and maximum value of national heterogeneity have been recalculated: new calculations solely base on the groups in which CEE participants were engaging with participants from WE and/or SE member states.

Table E.3: Network Densities of Consideration (Gender)

Group ID	Within-group references		Between-group references		Percentage point difference between within-and between-group	
	Men to Men	*Women to Women*	*Men to Women*	*Women to Men*	*Men*	*Women*
1	0.264	0.000	0.139	0.083	12.5	-8.3
2	0.286	0.107	0.125	0.143	16.1	-3.6
3	0.125	0.200	0.200	0.067	-7.5	13.3
4	0.304	0.100	0.208	0.229	9.6	-12.9
9	0.429	0.200	0.238	0.429	19.1	-22.9
10	0.214	0.119	0.204	0.143	1.0	-2.4
16	0.181	0.583	0.306	0.250	-12.5	33.3
17	0.179	0.119	0.143	0.143	3.6	-2.4
19	0.167	0.119	0.159	0.127	0.8	-0.8
20	0.119	0.067	0.024	0.071	9.5	-0.4
21	0.300	0.139	0.244	0.200	5.6	-6.1
23	0.222	0.250	0.111	0.267	11.1	-1.7
24	0.300	0.200	0.194	0.222	10.6	-2.2

Note: Entries are to be read as percentages. In group 21, for example, 30% of all possible references between men were made. In the same group, men referenced women in 24.4% of possible cases. This results in a slight tendency for homophily (5.6 percentage point difference). The last two columns are the entries of Figure 3.3.

Table E.4: Network Densities of Consideration (Working-class)

Group ID	Within-group references		Between-group references		Percentage point difference between within-and between-group	
	NW to NW	*W to W*	*NW to W*	*W to NW*	*NW*	*W*
1	0.155	0.000	0.182	0.273	-2.7	-27.3
2	0.220	0.000	0.028	0.083	19.2	-8.3
3	0.152	0.000	0.083	0.125	6.9	-12.5
4	---	---	---	---	---	---
9	0.417	0.250	0.250	0.278	16.7	-2.8
10	0.173	0.000	0.242	0.121	-6.9	-12.1
16	0.378	0.000	0.100	0.100	27.8	-10
17	0.189	0.000	0.056	0.111	13.3	-11.1
19	0.200	0.150	0.109	0.073	9.1	7.7
20	0.054	0.050	0.075	0.100	-2.1	-5.0
21	---	---	---	---	---	---
23	0.179	0.267	0.188	0.229	-0.9	3.8
24	0.300	0.000	0.050	0.100	25.0	-10

Note: See note in Table E.3 for an example. *NW*=Non working-class participants; *W*=working-class participants. Due to the lack of working-class participants in group 4 (only one *W*) and group 21 (no *W*), no consideration densities were calculated for the respective small group discussions. The last two columns are the entries of Figure 3.3.

Table E.5: Network Densities of Consideration (New CEE Members)

Group ID	Within-group references		Between-group references		Percentage point difference between within-and between-group	
	old to old	*new to new*	*old to new*	*new to old*	*old*	*new*
1	0.167	0.250	0.136	0.194	3.1	5.6
2	0.179	0.071	0.179	0.179	0.0	-10.8
9	0.690	0.167	0.190	0.238	50.0	-7.1
20	0.089	0.167	0.033	0.033	5.6	13.4
23	0.429	0.095	0.143	0.184	28.6	-8.9

Note: See note in Table E.3 for an example. *new*=participants from a new CEE member state; *old*=participants from an older member state (SE and/or WE). The last two columns are the entries of Figure 3.3.

Appendix F

ADDITIONAL ANALYSES (CHAPTER 4)

Bayesian inference was used to estimate the full model in order to identify irrelevant control variables (see section 4.4.4). In addition, Bayesian analysis provides a more intuitive understanding of the results in that it models the uncertainty of the estimated parameters given a particular sample. Since we deal with a non-randomly selected subsample of the Europolis group discussions, Bayesian inference may be appropriate (see, e.g., Hangartner et al., 2007). The Bayesian analysis was run in R using JAGS (Plummer, 2011). For both models, a two-level hierarchical model structure with varying intercepts was chosen. Following common practice, we start with non-informative prior distributions (see Gelman and Hill, 2007: 354). All independent variables are assumed to have a normal distribution. The precision follows a gamma distribution.[51]

For the item of regularization of irregular immigrants, two chains were run with 60'000 iterations, whereof the first 20'000 iterations were considered as burn-in. For the item of Europeanization of immigration, 80'000 iterations were run (20'000 burn-in). Both visual and statistical diagnostics were conducted. The trace plots show that the chains were mixing well and the density plots of the parameters approximate a normal distribution. We also ran several statistical diagnostics to test for non-convergence: The Gelman and Rubin (1992) and the Brooks and Gelman (1998) diagnostics both show values of 1, indicating that non-convergence was not an issue, neither within nor between chains. In addition, the Geweke (1991) diagnostics did not display excessive z-scores 2 or larger, indicating that the mean estimates converged from the first 10% to the last 50% of the Markov chains. In addition, the Heidelberger and Welch (1981) test indicates that the Markov chains are stationary for all parameters.

The posterior distributions in Table F.1 show that the predictors of age, belonging to a SE or CEE member state, interest in politics, and working-class are afflicted with a very high degree of uncertainty and may therefore be excluded from further analysis. Table F.2 shows the posterior distribution for the predictors of the Europeanization item. Again, interest in politics, working-class and country region show high degrees of uncertainty. The same accounts for education.

[51] Specification in JAGS: β ~ dnorm(0, .0001) and τ_y ~ dgamma(.001, .001) for the individual level; γ~ dnorm(0, .0001) and τ_α ~ dgamma(.01, .01) for the group level; μ_α ~ dnorm(0, .0001) for the random intercept mean.

Figure F.1: Posterior Distributions – Individual Level Predictors of Opinions on „Regularization of Irregular Immigrants“ (T3)

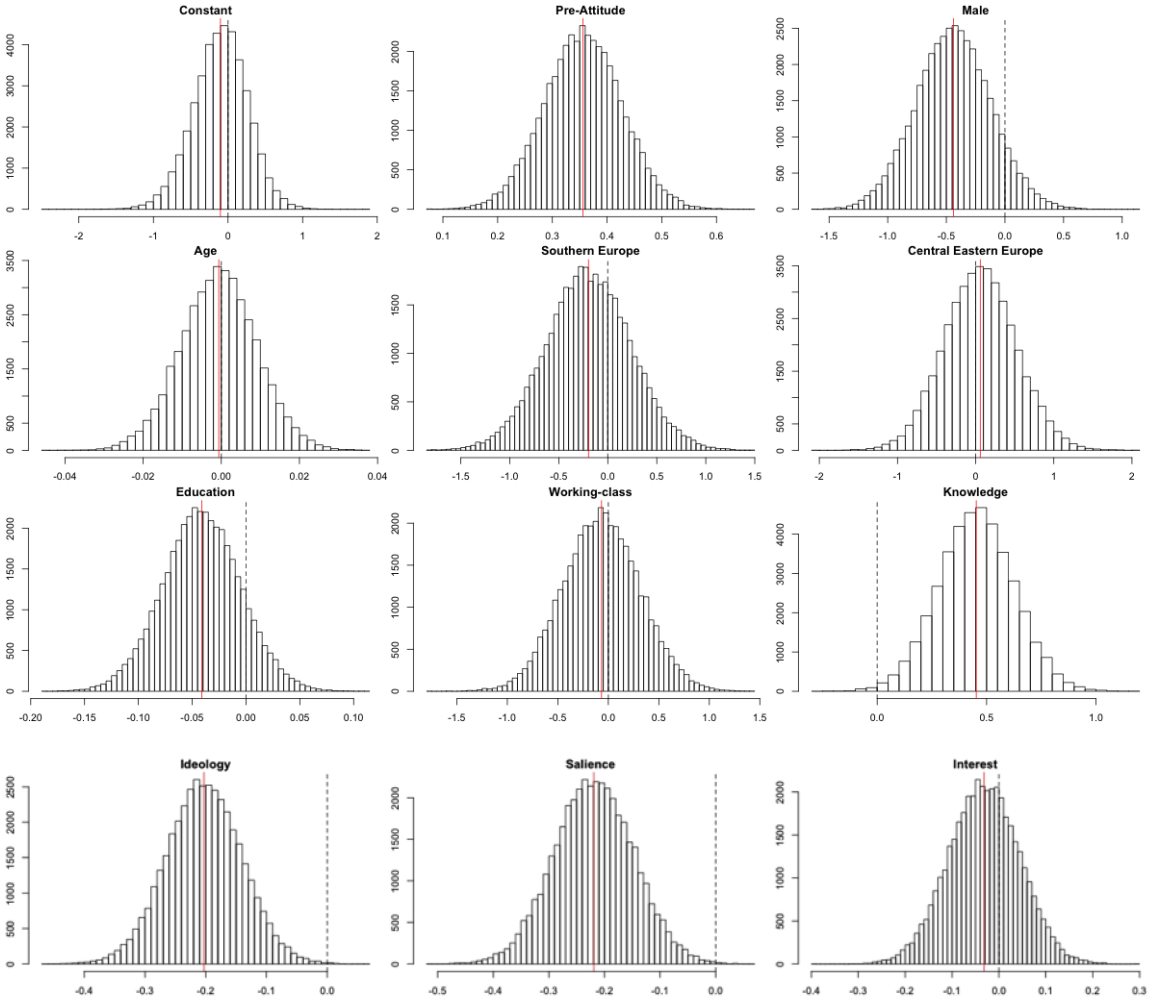
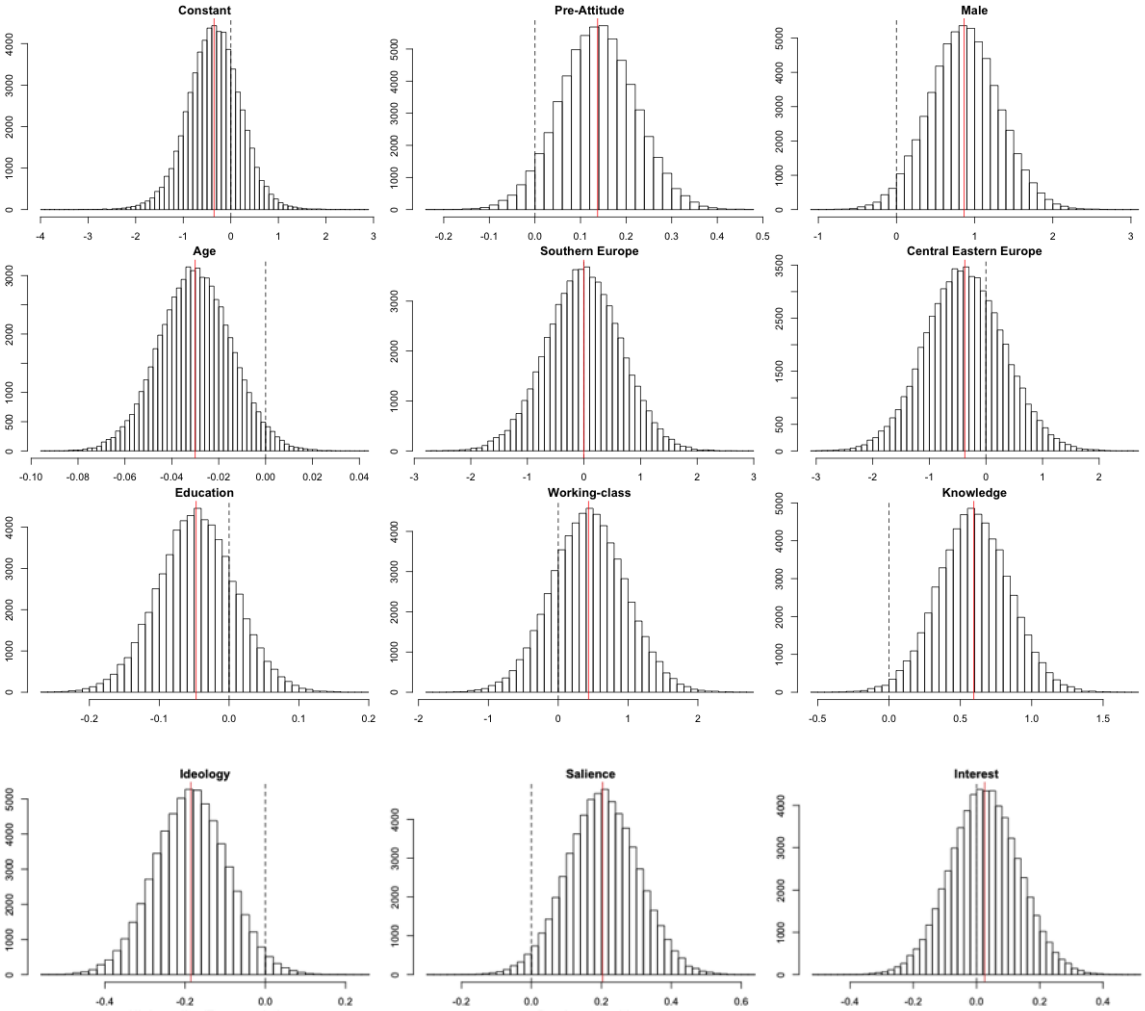


Figure F.2: Posterior Distributions – Individual Level Predictors of Opinions on „European Involvement in Immigration Affairs“ (T3)



Appendix G

CODING EXAMPLE DELIBERATIVE PERSUASION

Coding Example (European Involvement in Immigration Affairs)

> *A:* And if we speak of selection: in fact, it always lies in the interest of the state. Each state has its specific culture and specific values that need to be defended. And I think that every state should possess the autonomy to define criteria for controlling a problematic situation. (…) *(own translation from French)*

In this speech, speaker A argues against European involvement in immigration affairs. In his view, immigration requires accommodation of a host country's culture and values. Given that each EU-member state has its own culture, the right to decide on immigration issues should be left to each individual state. Due to the clear link between reason and conclusions, this speech received a (2) for qualified justification. A bit later, speaker B countered with the following statement:

> *B:* I heard that the first two speeches that were provided were in favor of leaving the decision on immigration to each individual EU-member state. (…) But in fact it is the European Union that needs to define the rules. Otherwise, we [the states] will do it because we know that Europe does not speak with one voice. There are four or five strong voices and various feeble ones. We have several voices and if we do not manage to overcome this state, we won't get anywhere. Just to offer you my opinion. I am not in favor of a federal state of Europe, no. But there are certain policies that need to be defined by the European Union, in a global sense. And immigration is one of these policies because we have internal and external borders. And we need to defend ourselves at the external borders. From the moment a person enters the European Union, she can go anywhere in the EU, to different states, or at least to the states that are part of the Schengen area. All EU-member states need to define common positions in respect of immigration policy. Since our internal borders are open, I do not see our future in acting independently from each other. If a person manages to enter the EU, she is already in the EU. *(own translation from French)*

Considering immigration as a global phenomenon, speaker B views the Europeanization of immigration policy as a logical consequence of open internal borders. Contrary to speaker A, B provides a more in-depth and differentiated account of the topic by outlining the differences between internal and external European borders. Therefore, the statement received the code (3) for sophisticated justification. In this group, not a single statement against European involvement had a comparable degree of argumentative sophistication. Consequently, the group receives a (1) for the group level variable indicating that the arguments in favor of European involvement were of higher justification rationality.

REFERENCES

Ackerman B and Fishkin JS (2002) Deliberation Day. *Journal of Political Philosophy*, 10(2), 129–152.

Alvarez RM and Brehm J (2002) *Hard Choices, Easy Answers: Values, Information and American Public Opinion*. Princeton, Princeton University Press.

Andersen VN and Hansen KM (2007) How Deliberation Makes Better Citizens: The Danish Deliberative Poll on the Euro. *European Journal of Political Research*, 46(4), 531–556.

Arkes HR and Blumer C (1984) Repetition and the Perceived Validity of Political Statements. In: *Midwestern Psychological Association*, Chicago, May 1984.

Azmanova A (2010) Deliberative Conflict and 'The Better Argument' Mystique. *The Good Society*, 19(1), 48–54.

Baccaro L, Craden C and Deville-Naggay M (2011) Should Foreigners Vote? Outcomes of a Deliberation Experiment in the City of Rousseau. In: *ECPR Joint Sessions of Workshops*, St. Gallen, CH, 12–16 April 2011.

Bächtiger A (2005) *The Real World of Deliberation: A Comparative Study of its Favorable Conditions in Legislatures*. Bern, Haupt.

Bächtiger A and Gerber M (2012) 'Gentlemanly Conversation' or Vigorous Contestation? An Exploratory Analysis of Communication Modes in a Transnational Deliberative Poll (Europolis). Unpublished Manuscript.

Bächtiger A and Hangartner D (2010) When Deliberative Theory Meets Empirical Political Science: Theoretical and Methodological Challenges in Political Deliberation. *Political Studies*, 58(4), 609–629.

Bächtiger A, Niemeyer S, Neblo M, et al. (2010a) Disentangling Diversity in Deliberative Democracy: Competing Theories, Their Blind Spots and Complementarities. *Journal of Political Philosophy*, 18(1), 32–63.

Bächtiger A, Pedrini S and Ryser M (2010b) Prozessanalyse politischer Entscheidungen: Deliberative Standards, Diskurstypen und Sequenzialisierung. In: Behnke J, Bräuninger T, and Shikano S (eds), *Jahrbuch für Handlungs- und Entscheidungstheorie*, Wiesbaden, VS Verlag für Sozialwissenschaften, pp. 193–226.

Bara J, Weale A and Bicquelet A (2007) Analysing Parliamentary Debate with Computer Assistance. *Swiss Political Science Review*, 13(4), 577–605.

Barabas J (2004) How Deliberation Affects Policy Opinions. *American Political Science Review*, 98(4), 687–701.

Bartle J (2000) Political Awareness, Opinion Constraint and the Stability of Ideological Positions. *Political Studies*, 48(3), 467–484.

Bates D, Maechler M and Bolker B (2011) Linear Mixed-Effects Models Using S4 Classes. Available from: http://lme4.r-forge.r-project.org/ (accessed 21 March 2012).

Benhabib S (1996) Toward a Deliberative Model of Democratic Legitimacy. In: Benhabib S (ed.), *Democracy and Difference: Contesting the Boundaries of the Political*, Cambridge University Press, pp. 67–94.

Benhabib S (2002) *The Claims of Culture. Equality and Diversity in the Global Era*. Princeton, University Press.

Bernstein B (1971-1990) *Class, Codes and Control: Theoretical Studies Towards a Sociology of Language.* 4 Volumes. London, Routledge & Kegan.

Bevir M and Ansari N (2012) Should Deliberative Democrats Eschew Modernist Social Science? Available from: http://wpsa.research.pdx.edu/meet/2012/ansari.pdf. (accessed 20 March 2012).

Bickford S (1996) *The Dissonance of Democracy: Listening, Conflict, and Citizenship*. New York, Cornell University Press.

Black L, Burkhalter S, Gastil J, et al. (2010) Methods for Analyzing and Measuring Group Deliberation. In: Bucy E and Holbert LR (eds), *Sourcebook for Political Communication Research: Methods, Measures, and Analytical Techniques*, New York, Routledge, pp. 323–345.

Bohman J (1997) Deliberative Democracy and Effective Social Freedom: Capabilities, Resources, and Opportunities. In: Bohman J and Rehg W (eds), *Deliberative Democracy: Essays on Reason and Politics*, Cambridge, MA, The MIT Press, pp. 321–248.

Bohman J (1998) Survey Article: The Coming of Age of Deliberative Democracy. *Journal of Political Philosophy*, 6(4), 400–425.

Bohman J (2006) Deliberative Democracy and the Epistemic Benefits of Diversity. *Episteme: A Journal of Social Epistemology*, 3(3), 175–191.

Bolker B (2010) *Worked Analysis of Owl Data*. Available from:
http://glmm.wdfiles.com/local--files/examples/Owls.pdf (accessed 4 March 2013).

Borgatti SP (2005) Centrality and Network Flow. *Social Networks*, 27(1), 55–71.

Borgatti SP, Everett M and Freeman L (2002) *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA, Analytic Technologies.

Brewer PR and Gross K (2005) Values, Framing, and Citizens' Thoughts about Policy Issues: Effects on Content and Quantity. *Political Psychology*, 26(6), 929–948.

Brooks SP and Gelman A (1998) General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434–455.

Brown MB (2006) Survey Article: Citizen Panels and the Concept of Representation. *The Journal of Political Philosophy*, 14(2), 203–225.

Browne WJ and Draper D (2006) A Comparison of Bayesian and Likelihood-Based Methods for Fitting Multilevel Models. *Bayesian Analysis*, 1(3), 473–514.

Bryan FM (1999) Direct Democracy and Civic Competence: The Case of Town Meeting. In: Elkin SL and Soltan KE (eds) *Citizen Competence and Democratic Institutions*, University Park, Pennsylvania State University Press, pp. 195–223.

Burkhalter S, Gastil J and Kelshaw T (2002) A Conceptual Definition and Theoretical Model of Public Deliberation in Small Face-to-Face Groups. *Communication Theory*, 12(4), 398–422.

Caluwaerts D (2012) Confrontation and Communication: Experiments on Deliberative Democracy in Linguistically Divided Belgium. Brussel, Vrije Universiteit Brussel.

Calvert A and Warren ME (2012) Can Deliberative Mini-Publics Overcome Anti-Deliberative Framing Effects? In: *Annual Meeting of the Western Political Science Association*,

Portland, OR, 22–24 March 2012.

Cappella JN, 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), 73–93.

Chambers S (1996) *Reasonable Democracy: Jürgen Habermas and the Politics of Discourse*. Ithaca, Cornell University Press.

Chambers S (2003) Deliberative Democratic Theory. *Annual Review of Political Science*, 6(1), 307–326.

Chambers S (2004) Behind Closed Doors: Publicity, Secrecy, and the Quality of Deliberation. *Journal of Political Philosophy*, 12(4), 389–410.

Chambers S (2009) Rhetoric and the Public Sphere: Has Deliberative Democracy Abandoned Mass Democracy? *Political Theory*, 37(3), 323–350.

Childs S (2004) A Feminised Style of Politics? Women MPs in the House of Commons. *The British Journal of Politics & International Relations*, 6(1), 3–19.

Chong D and Druckman JN (2007) Framing Public Opinion in Competitive Democracies. *American Political Science Review*, 101(4), 637–655.

Chong D and Druckman JN (2013) Counterframing Effects. *The Journal of Politics*, 75(1), 1–16.

Christiano T (2008) *The Constitution of Equality : Democratic Authority and Its Limits: Democratic Authority and Its Limits*. New York, Oxford University Press.

Cinalli M and O'Flynn I (2013) Public Deliberation, Network Analysis and the Political Integration of Muslims in Britain. *The British Journal of Politics & International Relations*, Early View.

Cohen Jacob (1960) A Coefficient of Agreement for Nominal Scales. *Educational and psychological measurement*, 20, 37–46.

Cohen Joshua (1989) Deliberation and Democratic Legitimacy. In: Hamlin A and Pettit P (eds), *The Good Polity*, Oxford, Blackwell, pp. 17–34.

Cohen Joshua (1996) Procedure and Substance in Deliberative Democracy. In: Benhabib S (ed.), *Democracy and Difference: Contesting the Boundaries of the Political*, Princeton, Princeton University Press, pp. 95–119.

Conover PJ and Searing DD (2005) Studying 'Everyday Political Talk' in the Deliberative System. *Acta Politica*, 40(3), 269–283.

Conover PJ, Searing DD and Crewe IM (2002) The Deliberative Potential of Political Discussion. *British Journal of Political Science*, 32(1), 21–62.

Converse PE (1964) The Nature of Belief Systems Among Mass Publics. In: Apter D (ed.), *Ideology and Discontent*, New York, Free Press, pp. 206–261.

Cook FL, Delli Carpini MX and Jacobs LR (2007) Who Deliberates? Discursive Participation in America. In: Rosenberg SW (ed.), *Can the People Govern? Deliberation, Participation and Democracy*, New York, Palgrave Macmillan, pp. 25–44.

Dahl RA (1989) *Democracy and its Critics*. New Haven, Yale University Press.

Delli Carpini MX, Cook FL and Jacobs LR (2004) Public Deliberations, Discursive

Participation and Citizen Engagement: A Review of the Empirical Literature. *Annual Review of Political Science*, 7(1), 315–344.

Delli Carpini MX and Keeter S (1996) *What Americans Know About Politics and Why It Matters*. London, Yale University Press.

Devine DJ, Clayton LD, Dunford BB, et al. (2001) Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups. *Psychology, Public Policy, and Law*, 7(3), 622–727.

Dickson ES, Hafer C and Landa D (2008) Cognition and Strategy: A Deliberation Experiment. *The Journal of Politics*, 70(4), 974–989.

Dobson A (2010) Democracy and Nature: Speaking and Listening. *Political Studies*, 58(4), 752–768.

Dobson A (2012) Listening: The New Democratic Deficit. *Political Studies*, 60(4), 843–859.

Doerr N (2009) Activists Beyond Language Borders? Multilingual Deliberative Democracy Experiments at the European Social Formus. In: *5th ECPR General Conference*, Potsdam, 10–12 September 2009.

Druckman JN (2001) The Implications of Framing Effects for Citizen Competence. *Political Behavior*, 23(3), 225–256.

Druckman JN and Nelson KR (2003) Framing and Deliberation: How Citizens' Conversations Limit Elite Influence. *American Journal of Political Science*, 47(4), 729–745.

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

Dryzek JS (2007) Theory, Evidence, and the Task of Deliberation. In: Rosenberg SW (ed.), *Can the People Govern? Deliberation, Participation and Democracy*, New York, Palgrave Macmillan, pp. 237–250.

Dryzek JS (2009) The Australian Citizens' Parliament: A World First. *Journal of Public Deliberation*, 5(1), 1–7.

Dryzek JS (2010) *Foundations and Frontiers of Deliberative Governance*. Oxford, Oxford University Press.

Dryzek JS, Bächtiger A and Milewicz K (2011) Toward a Deliberative Global Citizens' Assembly. *Global Policy*, 2(1), 33–42.

Dryzek JS and Niemeyer S (2008) Discursive Representation. *American Political Science Review*, 102(4), 481–493.

S and Haegel F (2007) Avoiding or Accepting Conflict in Public Talk. *British Journal of Political Science*, 37(1), 1–22.

Dutwin D (2003) The Character of Deliberation: Equality, Argument, and the Formation of Public Opinion. *International Journal of Public Opinion Research*, 15(3), 239–264.

Eliasoph N (1998) *Avoiding Politics: How Americans Produce Apathy in Everyday Life*. Cambridge, UK, Cambridge University Press.

Elster J (1997) The Market and the Forum: Three Varieties of Political Theory. In: Bohman J and Rehg W (eds), *Deliberative Democracy: Essays on Reason and Politics*, Cambridge,

MA, MIT Press, pp. 3–34.

Elster J (1998) Introduction. In: Elster J (ed.), *Deliberative Democracy*, Cambridge University Press, pp. 1–18.

Esterling KM (2011) 'Deliberative Disagreement' in U.S. Health Policy Committee Hearings. *Legislative Studies Quarterly*, 36(2), 169–198.

Estlund D (1997) Beyond Fairness and Deliberation: The Epistemic Dimension of Deliberative Authority. In: Bohman J and Rehg W (eds), *Deliberative Democracy: Essays on Reason and Politics*, Cambridge, MIT Press, pp. 127–160.

Estlund D (2000) Political Quality. *Social Philosophy and Policy*, 17(1), 127–160.

Estlund D (2009) Debate: On Christiano's The Constitution of Equality. *Journal of Political Philosophy*, 17(2), 241–252.

Farrar C, Fishkin JS, Green DP, et al. (2010) Disaggregating Deliberation's Effects: An Experiment Within a Deliberative Poll. *British Journal of Political Science*, 40(2), 333–347.

Fearon JD and Laitin DD (2003) Ethnicity, Insurgency, and Civil War. *American Political Science Review*, 97(1), 75–90.

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

Fishkin JS (2009) *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford, Oxford University Press.

Fishkin JS (2010) Response to Critics of When the People Speak: The Deliberative Deficit and What To Do About It. *The Good Society*, 19(1), 68–76.

Fishkin JS, He B, Luskin RC, et al. (2010) Deliberative Democracy in an Unlikely Place: Deliberative Polling in China. *British Journal of Political Science*, 40(2), 435–448.

Fishkin JS and Luskin RC (2005) Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion. *Acta Politica*, 40(3), 284–298.

Fishkin JS and Luskin RC (2006) Broadcasts of Deliberative Polls: Aspirations and Effects. *British Journal of Political Science*, 36(1), 184–188.

Fraser N (1992) Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. In: Calhoun C (ed.), *Habermas and the Public Sphere*, Cambridge, MIT Press, pp. 109–142.

Freeman GP (1995) Modes of Immigration Politics in Liberal Democratic States. *International Migration Review*, 29(4), 881–902.

Freeman LC (1979) Centrality in Social Networks: Conceptual Clarification. *Social Networks*, 1(3), 215–239.

Fuchs D and Klingemann H-D (2002) Eastward Enlargement of the European Union and the Identity of Europe. *West European Politics*, 25(2), 19–54.

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

Gambetta D (1998) Claro! An Essay on Discursive Machismo. In: Elster John (ed.), *Deliberative Democracy*, Cambridge, Cambridge University Press, pp. 19–43.

Gastil J, Black L and Moscovitz K (2008) Ideology, Attitude Change, and Deliberation in Small Face-to-Face Groups. *Political Communication*, 25(1), 23–46.

Geddes A (2003) *The Politics of Migration and Immigration in Europe*. London, SAGE.

Gelman A and Hill J (2007) *Data Analysis Using Regression And Multilevel/Hierarchical Models*. New York, Cambridge University Press.

Gelman A and Rubin DB (1992) Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4), 457–472.

Gerber M (2011) Who are the Voices of Europe? Evidence from a pan-European Deliberative Poll. In: *ECPR General Conference*, Reykjavik, 25–27 August 2011.

Geweke J (1991) Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments. In: Bernado JM, Berger J, Dawid A, et al. (eds), *Bayesian Statistics 4*, Clarendon Press, Available from: http://minneapolisfed.org/Research/sr/sr148.pdf (accessed 13 September 2012).

Goodin RE (2000) Democratic Deliberation Within. *Philosophy & Public Affairs*, 29(1), 81–109.

Goodin RE and Dryzek JS (2006) Deliberative Impacts: The Macro-Political Uptake of Mini-Publics. *Politics & Society*, 34(2), 219 –244.

Goodin RE and Ratner SR (2011) Democratizing International Law. *Global Policy*, 2(3), 241–247.

Graham T and Witschge T (2003) In Search of Online Deliberation : Towards a New Method for Examining the Quality of Online Discussions. *Communications*, 28(2), 173–204.

Grönlund K, Setälä M and Herne K (2010) Deliberation and Civic Virtue: Lessons from a Citizen Deliberation Experiment. *European Political Science Review*, 2(1), 95–117.

Gutmann A and Thompson DF (1996) *Democracy and Disagreement*. Cambridge, Harvard University Press.

Gutmann A and Thompson D (2002) Deliberative Democracy Beyond Process. *Journal of Political Philosophy*, 10(2), 153–174.

Gutmann A and Thompson DF (2004) *Why Deliberative Democracy?* Princeton, Princeton University Press.

Habermas J (1991) *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Cambridge, MIT Press.

Habermas J (1996) *Between Facts and Norms*. Cambridge, MIT Press.

Habermas J (2005) Concluding Comments on Empirical Approaches to Deliberative Politics. *Acta Politica*, 40(2), 384–392.

Hangartner D, Bächtiger A, Grünenfelder R, et al. (2007) Mixing Habermas with Bayes: Methodological and Theoretical Advances in the Study of Deliberation. *Swiss Political Science Review*, 13(4), 607–644.

Hansen KM (2004) *Deliberative Democracy and Opinion Formation*. Odense, University Press of Southern Denmark.

Hansen KM (2007) The Sophisticated Public: The Effect of Competing Frames on Public Opinion. *Scandinavian Political Studies*, 30(3), 377–396.

Hansen KM (2010) The Equality Paradox of Deliberative Democracy: A Review of Theoretical Concepts and Empirical Findings. In: Wolf MR, Morales L, and Ikeda K (eds), *Political Discussions in Modern Democracies*, New York, Routledge, pp. 26–43.

Heidelberger P and Welch PD (1981) A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations. *Communications of the ACM*, 24(4), 233–245.

Hibbing JR and Theiss-Morse E (2002) *Stealth Democracy: Americans' Beliefs About How Government Should Work*. Cambridge, Cambridge University Press.

Himmelroos S (2011) Democratically Speaking: Can Citizen Deliberation Be Considered Fair and Equal? In: *ECPR Joint Sessions of Workshops*, St. Gallen, CH, 12–16 April 2011.

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

Holzinger K (2001) Kommunikationsmodi und Handlungstypen in den Internationalen Beziehungen. *Zeitschrift für Internationale Beziehungen*, 8(2), 243–286.

Ibarra H (1993) Network Centrality, Power, and Innovation Involvement: Determinants of Technical and Administrative Roles. *Academy of Management Journal*, 36(3), 471–501.

Ikuenobe P (2001) Questioning as an Epistemic Process of Critical Thinking. *Educational Philosophy and Theory*, 33(3-4), 325–341.

Isernia P, Fishkin JS, Steiner J, et al. (2013) Toward a European Public Sphere: The Europolis Project. In: Kies R and Nanz P (eds), *Is Europe Listening to Us? Sucesses and Failures of EU Citizen Consultation*, Aldershot, Ashgate Publishing. Forthcoming.

Jackman S (2001) Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference, and Modell Checking. *Political Analysis*, 9(3), 227–241.

Janssen D and Kies R (2005) Online Forums and Deliberative Democracy. *Acta Politica*, 40(3), 317–335.

Jennstal J (2012) *Traits and Talks: Lessons About Personality and Deliberation from the Negotiations Between Nelson Mandela & Fw De Klerk*. Uppsala, University of Uppsala.

Kanra B (2005) Democracy, Islam and Dialogue: The Case of Turkey. *Government and Opposition*, 40(4), 515–539.

Karp JA and Banducci SA (2007) Party Mobilization and Political Participation in New and Old Democracies. *Party Politics*, 13(2), 217–234.

Karpowitz CF and Mendelberg T (2007) Groups and Deliberation. *Swiss Political Science Review*, 13(4), 645–662.

Karpowitz CF, Mendelberg T and Shaker L (2012) Gender Inequality in Deliberative Participation. *American Political Science Review*, 106(3), 533–547.

Karpowitz CF, Raphael C and Hammond AS (2009) Deliberative Democracy and Inequality: Two Cheers for Enclave Deliberation among the Disempowered. *Politics & Society*, 37(4), 576–615.

Kim N, Siu A and Sood G (2010) Minority Report: Impact of Opinion Minorities on Deliberation. In: *Annual Meeting of the International Communication Association*, Singapore, 2011.

King M (2009) A Critical Assessment of Steenbergen et al's Discourse Quality Index. *Roundhouse*, 1(1), 1–8.

Knight J and Johnson J (1994) Aggregation and Deliberation: On the Possibility of Democratic Legitimacy. *Political Theory*, 22(2), 277–296.

Knight J and Johnson J (1997) What Sort of Political Equality Does Deliberative Democracy Require? In: Bohman J and Rehg W (eds), *Deliberative Democracy: Essays on Reason and Politics*, Cambridge, The MIT Press, pp. 279–319.

Krackhardt D (1990) Assessing the Political Landscape: Structure, Cognition, and Power in Organizations. *Administrative Science Quarterly*, 35(2), 342–369.

Krupnikov Y, Levine AS and Lupia A (2007) Is Deliberative Democracy Psychologically Plausible? Available at: http://www.allacademic.com/meta/p209389_index.html (accessed 23 December 2010).

Kymlicka W (2000) Nation-Building and Minority Rights: Comparing West and East. *Journal of Ethnic and Migration Studies*, 26(2), 183–212.

Landemore H and Mercier H (2012) Talking it Out with Others vs. Deliberation Within and the Law of Group Polarization: Some Implications of the Argumentative Theory of Reasoning. *Análise Social*, 205(4), 910–934.

Landis JR and Koch GG (1977) The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174.

Lavine HG, Johnston CD and Steenbergen MR (2012) *The Ambivalent Partisan: How Critical Loyalty Promotes Democracy*. New York, Oxford University Press.

Lazer D, Rubineau B, Chetkovich C, et al. (2010) The Coevolution of Networks and Political Attitudes. *Political Communication*, 27(3), 248–274.

Lord C and Tamvaki D (2013) The Politics of Justification? Applying the 'Discourse Quality Index' to the Study of the European Parliament. *European Political Science Review*, 5(1), 27–54.

Lupia A (2002) Deliberation Disconnected: What it Takes to Improve Civic Competence. *Law and Contemporary Problems*, 65(3), 133–150.

Lupia A and McCubbins MD (1998) *The Democratic Dilemma*. Cambridge, Cambridge University Press.

Luskin RC (2002) *True versus Measured Information Gain*. University of Texas at Austin. Available at: http://cdd.stanford.edu/research/papers/2001/true-infogain.pdf (accessed 10 December 2012).

Luskin RC, Fishkin JS and Jowell R (2002) Considered Opinions: Deliberative Polling in Britain. *British Journal of Political Science*, 32(3), 455–487.

Luskin RC, O'Flynn I, Fishkin JS, et al. (2013) Deliberating across Deep Divides. *Political Studies*. Early View.

Macedo S (1999) Introduction. In: Macedo S (ed.), *Deliberative Politics : Essays on Democracy and Disagreement*, New York, Oxford University Press, pp. 3–14.

Mansbridge J (1983) *Beyond Adversary Democracy*. Chicago, IL, University of Chicago Press.

Mansbridge J (1996a) Reconstructing Democracy. In: Hirschmann NJ and Di Stefano C (eds), *Revisioning the Political: Feminist Reconstructions of Traditional Concepts in Western Political Theory*, New York, Westview, pp. 117–158.

Mansbridge J (1996b) Using Power/Fighting Power: The Polity. In: Benhabib S (ed.), *Democracy and Difference: Contesting Boundaries of the Political*, Princeton, Princeton University Press, pp. 46–66.

Mansbridge J (1999) Everyday Talk in the Deliberative System. In: Macedo S (ed.), *Deliberative Politics: Essays on Democracy and Disagreement*, New York, Oxford University Press, pp. 211–239.

Mansbridge J (2007) Deliberative Democracy‘ or ’Democratic Deliberation. In: Rosenberg SW (ed.), *Can the People Govern? Deliberation, Participation and Democracy*, New York, Palgrave Macmillan, pp. 251–271.

Mansbridge J (2010) Deliberative Polling as the Gold Standard. *The Good Society*, 19(1), 55–62.

Mansbridge J, Bohman J, Chambers S, et al. (2012) A Systemic Approach to Deliberative Democracy. In: Parkinson J and Mansbridge J (eds), *Deliberative Systems: Deliberative Democracy at the Large Scale*, Cambridge, Cambridge University Press, pp. 1–26.

Mansbridge J, Bohman J, Chambers S, et al. (2010) The Place of Self-Interest and the Role of Power in Deliberative Democracy*. *Journal of Political Philosophy*, 18(1), 64–100.

Mansbridge J, Hartz-Karp J, Amengual M, et al. (2006) Norms of Deliberation: An Inductive Study. *Journal of Public Deliberation*, 1(2), 1–47.

McCrae RR and Costa PT (1987) Validation of the Five-Factor Model of Personality across Instruments and Observers. *Journal of Personality and Social Psychology*, 52(1), 81–90.

McPherson M, Smith-Lovin L and Cook JM (2001) Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1), 415–444.

Mendelberg T (2002) The Deliberative Citizen: Theory and Evidence. In: Delli Carpini MX, Huddy L, and Shapiro RY (eds), *Political Decision Making, Deliberation and Participation: Research in Micropolitics, Volume 6*, Greenwich, CT, JAI Press, pp. 151–193.

Mendelberg T (2005) Bringing the Group Back Into Political Psychology: Erik H. Erikson Early Career Award Address. *Political Psychology*, 26(4), 637–650.

Mendelberg T and Oleske J (2000) Race and Public Deliberation. *Political Communication*, 17(2), 169–191.

Mercier H and Landemore H (2012) Reasoning Is for Arguing: Understanding the Successes and Failures of Deliberation. *Political Psychology*, 33(2), 243–258.

Min S (2009) Deliberation, East Meets West: Exploring the Cultural Dimension of Citizen Deliberation. *Acta Politica*, 44(4), 439–458.

Mondak JJ, Hibbing MV, Canache D, et al. (2010) Personality and Civic Engagement: An Integrative Framework for the Study of Trait Effects on Political Behavior. *American Political Science Review*, 104(1), 85–110.

Moons WG, Mackie DM and Garcia-Marques T (2009) The Impact of Repetition-Induced Familiarity on Agreement with Weak and Strong Arguments. *Journal of Personality and*

Social Psychology, 96(1), 32–44.

Moore A (2012) Following From the Front: Theorizing Deliberative Facilitation. *Critical Policy Studies*, 6(2), 146–162.

Morrell ME (2010) *Empathy and Democracy: Feeling, Thinking, and Deliberation*. University Park, Pennsylvania State University Press.

Mucciaroni G and Quirk PJ (2006) *Deliberative Choices: Debating Public Policy in Congress*. Chicago, University of Chicago Press.

Mucciaroni G and Quirk PJ (2010) Rhetoric and Reality: Going Beyond Discourse Ethics in Assessing Legislative Deliberation. *Legisprudence*, 4(1), 35–52.

Muhlberger P (2007) *Report to the Deliberative Democracy Consortium: Building a Deliberation Measurement Toolbox. Version 2; 12/06/2007*. Available from: www.geocities.com/pmuhl78/DDCReport.pdf (accessed 11 September 2012).

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

Mutz DC (2008) Is Deliberative Democracy a Falsifiable Theory? *Annual Review of Political Science*, 11(1), 521–538.

Myers DC (2012) Interests, Information and Minority Influence in Political Deliberation. In: *MPSA Conference in Chicago,* April 11-15 2012.

Naurin D (2010) Most Common When Least Important: Deliberation in the European Union Council of Ministers. *British Journal of Political Science*, 40(1), 31–50.

Neblo MA (2007a) Change for the Better? Linking the Mechanisms of Deliberative Opinion Change to Normative Theory. Available at: polisci.osu.edu/faculty/mneblo/papers/ChangeC4.pdf (accessed 8 January 2012).

Neblo M (2007b) Family Disputes: Diversity in Defining and Measuring Deliberation. *Swiss Political Science Review*, 13(4), 527–557.

Neblo MA, Esterling KM, Kennedy RP, et al. (2010) Who Wants To Deliberate—And Why? *American Political Science Review*, 104(3), 566–583.

Nelson TE, Clawson RA and Oxley ZM (1997) Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance. *The American Political Science Review*, 91(3), 567–583.

Nie NH, Junn J and Stehlik-Barry K (1996) *Education and Democratic Citizenship in America*. Chicago, University of Chicago Press.

Norris P (1996) Women Politicians: Transforming Westminster? *Parliamentary Affairs*, 49(1), 89–102.

Olsen ED and Trenz H-J (2013) From Citizens' Deliberation to Popular Will Formation? Generating Democratic Legitimacy in Transnational Deliberative Polling. *Political Studies*. Early View.

Papacharissi Z (2004) Democracy Online: Civility, Politeness, and the Democratic Potential of Online Political Discussion Groups. *New Media & Society*, 6(2), 259–283.

Parkinson J (2003) Legitimacy Problems in Deliberative Democracy. *Political Studies*, 51(1), 180–196.

Parkinson J (2006a) *Deliberating in the Real World: Problems of Legitimacy in Deliberative*

Democracy. Oxford, Oxford University Press.

Parkinson J (2006b) Of Scale and Straw Men: A Reply to Fishkin and Luskin. *British Journal of Political Science*, 36(1), 189–191.

Parkinson J (2006c) Rickety Bridges: Using the Media in Deliberative Democracy. *British Journal of Political Science*, 36(1), 175–183.

Parkinson J (2012) *Democracy and Public Space: The Physical Sites of Democratic Performance*. Oxford, Oxford University Press.

Pedrini S, Bächtiger A and Steenbergen MR (2013) Deliberative Inclusion of Minorities: Patterns of Reciprocity Among Linguistic Groups in Switzerland. *European Political Science Review*. Early View.

Petty RE and Cacioppo JT (1984) The Effects of Involvement on Responses to Argument Quantity and Quality: Central and Peripheral Routes to Persuasion. *Journal of Personality and Social Psychology*, 46(1), 69–81.

Petty RE, Cacioppo JT and Schumann D (1983) Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement. *Journal of Consumer Research*, 10(2), 135–146.

Plomin R, McClearn G, Smith D, et al. (1994) DNA Markers Associated with High Versus Low IQ: The IQ Quantitative Trait Loci (QTL) Project. *Behavior Genetics*, 24(2), 107–118.

Plummer M (2011) Bayesian Graphical Models Using MCMC. Available from: http://mcmc-jags.sourceforge.net (accessed 14 September 2012).

Polletta F and Lee J (2006) Is Telling Stories Good for Democracy? Rhetoric in Public Deliberation after 9/11. *American Sociological Review*, 71(5), 699–723.

Posner RA (2003) *Law, Pragmatism, and Democracy*. Cambridge, MA, Harvard University Press.

Posner R (2004) Smooth Sailing. *Legal Affairs*, Available from: http://www.legalaffairs.org/issues/January-February-2004/feature_posner_janfeb04.msp. (accessed: 9 May 2011).

Price V (2009) Citizens Deliberating Online: Theory and Some Evidence. In: Davies T and Gangadharan SP (eds), *Online Deliberation: Design, Research, and Practice*, Stanford, CSLI Publications, pp. 37–58.

Price V, Cappella JN and Nir L (2002) Does Disagreement Contribute to More Deliberative Opinion? *Political Communication*, 19(1), 95–112.

Price V and Neijens P (1998) Deliberative Polls: Toward Improved Measures of 'Informed' Public Opinion? *International Journal of Public Opinion Research*, 10(2), 145 –176.

Price V, Nir L and Cappella JN (2005) Framing Public Discussion of Gay Civil Unions. *Public Opinion Quarterly*, 69(2), 179–212.

Rao V and Sanyal P (2009) *Dignity through Discourse: Poverty and the Culture of Deliberation in Indian Village Democracies*. Policy Research working paper, The World Bank, Development Research Gruop, Poverty Team.

Rasch G (1980) *Probabilistic Models for Intelligence and Attainment Tests*. Chicago, IL,

University of Chicago Press.

Ratner SR (2008) Communicative Rationality in the Citizens' Assembly and Referendum Processes. In: Warren ME and Pearse H (eds), *Designing Deliberative Democracy: The British Columbia Citizens' Assembly*, Cambridge, UK, Cambridge University Press, pp. 145–165.

Rawls J (1971) *A Theory of Justice*. Cambridge, MA, The Belknap Press.

Rawls J (1993) *Political Liberalism*. New York, Columbia University Press.

Reise SP, Ainsworth AT and Haviland MG (2005) Item Response Theory Fundamentals, Applications, and Promise in Psychological Research. *Current Directions in Psychological Science*, 14(2), 95–101.

Reykowski J (2006) Deliberative Democracy and 'Human Nature': An Empirical Approach. *Political Psychology*, 27(3), 323–346.

Rosenberg SW (2005) The Empirical Study of Deliberative Democracy: Setting a Research Agenda. *Acta Politica*, 40(2), 212–224.

Rosenberg SW (2007) Rethinking Democratic Deliberation: The Limits and Potential of Citizen Participation. *Polity*, 39(3), 335–360.

Ryfe DM (2005) Does Deliberative Democracy Work? *Annual Review of Political Science*, 8, 49–71.

Ryfe DM (2006) Narrative and Deliberation in Small Group Forums. *Journal of Applied Communication Research*, 34(1), 72–93.

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*, 42(3), 617–640.

Sanders LM (1997) Against Deliberation. *Political Theory*, 25(3), 347–376.

Sanders LM (2010) Making Deliberation Cooler. *The Good Society*, 19(1), 41–47.

Sass J (2006) Does Culture Count in Political Deliberation? Claro! In: *Conference on Dialogue Across Difference*, Canberra, 4-5 December 2006.

Saward M (2003) Enacting Democracy. *Political Studies*, 51(1), 161–179.

Schneiderhan E and Khan S (2008) Reasons and Inclusion: The Foundation of Deliberation. *Sociological Theory*, 26(1), 1–24.

Schulz-Hardt S, Frey D, Lüthgens C, et al. (2000) Biased Information Search in Group Decision Making. *Journal of Personality and Social Psychology*, 78(4), 655–669.

Siegel S (1956) *Nonparametric Statistics for the Behavioral Sciences*. New York, McGraw-Hill.

Siu A (2009) Look Who's Talking: Examining Social Influence, Opinion Change and Argument Quality in Deliberation. Stanford University.

Smith G (2009) *Democratic Innovations: Designing Institutions for Citizen Participation-Theories of Institutional Design*. New York, Cambridge University Press.

Smith G (2011) Survey Article Democratic Innovations: Bringing Theory and Practice into Dialogue. *Philosophy Compass*, 6(12), 895–901.

Smith G and Wales C (2000) Citizens' Juries and Deliberative Democracy. *Political Studies*,

48(1), 51–65.

Sniderman PM and Theriault SM (2004) The Structure of Political Argument and the Logic of Issue Framing. In: Saris WE and Sniderman PM (eds), *Studies in Public Opinion*, Princeton, Princeton University Press, pp. 133–165.

Spörndli M (2004) *Diskurs und Entscheidung: eine empirische Analyse kommunikativen Handelns im deutschen Vermittlungsausschuss*. Wiesbaden, VS Verlag.

Steenbergen MR, Bächtiger A, Spörndli M, et al. (2003) Measuring Political Deliberation : A Discourse Quality Index. *Comparative European Politics*, 1(1), 21–48.

Steenbergen MR and Jones BS (2002) Modeling Multilevel Data Structures. *American Journal of Political Science*, 46(1), 218–237.

Stegmueller D (2013) How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches. *American Journal of Political Science*. Early View.

Steiner J (2012) *The Foundations of Deliberative Democracy: Empirical Research and Normative Implications*. Cambridge, Cambridge University Press.

Steiner J, Bächtiger A, Spörndli M, et al. (2004) *Deliberative Politics in Action*. Cambridge, Cambridge University Press.

Stromer-Galley J (2007) Measuring Deliberation Content : A Coding Scheme. *Journal of Public Deliberation*, 3(1), 1–35.

Sturgis P (2003) Knowledge and Collective Preferences. *Sociological Methods & Research*, 31(4), 453 –485.

Sunstein CR (2002) The Law of Group Polarization. *Journal of Political Philosophy*, 10(2), 175–195.

Thompson DF (2008) Deliberative Democratic Theory and Empirical Political Science. *Annual Review of Political Science*, 11(1), 497–520.

Treier S and Jackman S (2008) Democracy as a Latent Variable. *American Journal of Political Science*, 52(1), 201–217.

Trénel M (2009) Facilitation and Inclusive Deliberation. In: Davies T and Gangadharan SP (eds), *Online Deliberation: Design, Research, and Practice*, Stanford University, Center for the Study of Language and Information/SRI, pp. 253–257.

Ulbert C and Risse T (2005) Deliberately Changing the Discourse: What Does Make Arguing Effective? *Acta Politica*, 40(3), 351–367.

Warren ME (2001) *Democracy and Association*. Princeton, Princeton University Press.

Warren ME and Pearse H (2008) *Designing Deliberative Democracy: The British Columbia Citizens' Assembly*. Cambridge, Cambridge University Press.

Weale A, Bicquelet A and Bara J (2012) Debating Abortion, Deliberative Reciprocity and Parliamentary Advocacy. *Political Studies*, 60(3), 643–667.

Williams MS (2000) The Uneasy Alliance of Group Representation and Deliberative Democracy. In: Kymlicka W and Norman W (eds), *Citizenship in Diverse Societies*, Oxford, University Press, pp. 124–153.

Young IM (2002) *Inclusion and Democracy*. Paperback edition. Oxford, Oxford University

Press.

Zaller J (1992) *The Nature and Origins of Mass Opinion*. Cambridge, Cambridge University Press.

Zaller J and Feldman S (1992) A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences. *American Journal of Political Science*, 36(3), 579–616.

Zhang L and Huang J (2001) Thinking Styles and the Five-Factor Model of Personality. *European Journal of Personality*, 15(6), 465–476.

Selbständigkeitserklärung

„Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Koautorenschaften sowie alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetzes vom 5. September 1996 über die Universität zum Entzug des aufgrund dieser Arbeit verliehenen Titels berechtigt ist.“

Thun, 01. April 2013 Marlène Gerber