

This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

---

## Impact of Temporality and Identifiability in Online Deliberations on Discussion Quality: An Experimental Study

Strandberg, Kim; Berg, Jan-Anders

*Published in:*  
Javnost: the Public

*DOI:*  
[10.1080/13183222.2015.1041230](https://doi.org/10.1080/13183222.2015.1041230)

Published: 01/01/2015

[Link to publication](#)

*Please cite the original version:*  
Strandberg, K., & Berg, J-A. (2015). Impact of Temporality and Identifiability in Online Deliberations on Discussion Quality: An Experimental Study. *Javnost: the Public*, 22(2), 164–180.  
<https://doi.org/10.1080/13183222.2015.1041230>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This article was downloaded by: [Åbo Akademis Bibliotek]

On: 11 August 2015, At: 03:14

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London, SW1P 1WG



## Javnost - The Public: Journal of the European Institute for Communication and Culture

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/rjav20>

### Impact of Temporality and Identifiability in Online Deliberations on Discussion Quality: An Experimental Study

Kim Strandberg & Janne Berg

Published online: 15 Jul 2015.



CrossMark

[Click for updates](#)

To cite this article: Kim Strandberg & Janne Berg (2015) Impact of Temporality and Identifiability in Online Deliberations on Discussion Quality: An Experimental Study, Javnost - The Public: Journal of the European Institute for Communication and Culture, 22:2, 164-180, DOI: [10.1080/13183222.2015.1041230](https://doi.org/10.1080/13183222.2015.1041230)

To link to this article: <http://dx.doi.org/10.1080/13183222.2015.1041230>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

# IMPACT OF TEMPORALITY AND IDENTIFIABILITY IN ONLINE DELIBERATIONS ON DISCUSSION QUALITY: AN EXPERIMENTAL STUDY

**Kim Strandberg and Janne Berg**

*There is a perception that citizen deliberation brings about higher-quality discussions than discussions where deliberative norms are not used. Often, deliberations are realised in mini-publics in which certain contextual features ensure, a priori, that the discussions are likely to be of a high quality. However, few studies have as yet explored the boundaries of deliberation; that is, contemplated what happens to discussion quality if the ideal-speech situation is strayed away from. To address this point, this article reports on an online experiment in which the discursive setting of citizen deliberations is manipulated. The experiment (n = 50 participants) was carried out online in Finland in November 2013 in order to test the impact on discussion quality related to two factors: the temporality (asynchronous or synchronous discussions) and identifiability of participants (known or anonymous) in an online deliberation. The findings clearly indicate that asynchronous discussions have the most positive influence on discussion quality. Moreover, the identifiability factor only had a small weak influence on discussion quality, and there was only one weak interaction effect between the two factors.*

**KEYWORDS** online deliberation; discussion quality; temporality; identifiability; Finland

## INTRODUCTION

Starting, perhaps, with Barber's (1984) notion of "strong democratic talk" and also Dahl's (1989, 169–179) vision of an inclusive democratic system, an increased interest in deliberative democracy has recently been evident within democratic theory. Drawing on the Habermasian vision of a public sphere in which lay citizens gather to rationally and critically discuss societal matters, it has been proposed that a democratic system based on the inclusion of citizens in political discussion, deliberation and decision-making increases the functioning and legitimacy of the political system (for example, Barber 1984, 133–135; Habermas 1996, 110). Thus, the core of deliberative theory is the deliberative process itself. This process, however, needs to meet certain normative criteria—for instance, inclusiveness, equality, reciprocity, rationality and mutual respect (for example, Dahl 1989, 169–179; Kies 2010, 40–53; Stromer-Galley 2007)—if it is to be considered a democratically beneficial discursive process.

However, as pointed out by Dahl (1989, 169–170), such ideal discursive processes are unlikely to appear by themselves. As such, there have been several scholarly efforts to convert the theoretical ideals into practice in so-called mini-publics, citizen juries, consensus conferences and deliberative polls (for example, Fishkin, Luskin, and Jowell 2000; Grönlund,

Setälä, and Herne 2010). Usually, these deliberation events employ discussion rules to facilitate and ensure that the discursive process adheres to the normative deliberative ideals (Ryfe 2005, 63). With the advent of the Internet, an array of deliberation events has been carried out online, mainly as scholarly experiments (overviews in, for example, Davies and Gangadharan 2009; Manosevitch 2014; Strandberg & Grönlund 2013; Towne and Herb- sleb 2012). Although many of both the offline and online deliberations have achieved discussions close to the deliberative ideals and found that these have had positive effects on the participating citizens (for example, Fishkin, Luskin, and Jowell 2000; Grönlund, Setälä, and Herne 2010), one might question whether such findings are actually surprising. As Coleman and Moss (2012, 6) remark, most studies have only applied “the social and technical conditions under which it [online deliberation] is most likely to be successful”. So, basically, most studies have been most-likely cases in which support for deliberative theory has been found. In terms of theory generation, little knowledge has been gained on the generalisability of deliberative theory. As such, using less-than-ideal preconditions and deliberative designs would arguably subject deliberative theory to a more critical test. Additionally, as remarked by Author’s own work (2014, 15–16), although many existing studies have looked at the democratic outcomes of online deliberation (i.e. how citizens are affected through deliberation), very few studies of online deliberation have actually measured the quality of the discussion (i.e. how citizens discuss) and whether this meets the procedural ideals of deliberative theory. Thus, little is yet known about the discussion quality of online deliberations.

In this article we report on the findings from an online lab-in-the-field experiment (Morton and Williams 2010, 296–297) designed to test the boundaries of deliberative theory. In simple terms we ask the question concerning what happens to discussion quality in an online deliberation if we stray from the ideal deliberative setting. To this end, we designed a discursive context, manipulating both the temporality of the discussion (asynchronous or synchronous) and the identifiability of participants (anonymous or known) in a  $2 \times 2$  basic factorial experimental design. As these two factors have been regarded as having a potential influence on discussions between citizens in general (Jensen 2003; Janssen and Kies 2005), and deliberation specifically (Bächtiger 2011), their inclusion in a deliberative discussion provides valuable knowledge on how external conditions affect the quality of deliberative processes, at least in an online context. Additionally, it is perceivable that the study provides some tentative insight into the “democratic value” of regular online discussions in as much as these are mostly conducted in less-than-ideal deliberative conditions.

## Theoretical Backdrop

### *Potential Impact of Temporality on Discussion Quality in Online Deliberations*

In a meta-analysis of experiments in online deliberation (cf. Strandberg and Grönlund 2013), it was found that both synchronous (for example, Luskin, Fishkin, and Iyengar 2006; Price and Capella 2002) and asynchronous discussions (for example, Albrecht 2006; Rhee and Kim 2009) have been applied rather evenly in existing studies. Although it is often assumed that asynchronous discussions are more conducive to deliberation than synchronous discussion—that is, through allowing time for reflection and contemplation

(Janssen and Kies 2005, 321; Smith, John, and Sturgis 2013, 710)—no studies have empirically tested whether this holds true in a systematic comparison with synchronous discussions. However, in reviewing the literature, some traces of associated thoughts concerning how temporality (asynchronous and synchronous online discussions) might be connected to discussion quality are discernible—albeit that they are very few. Nonetheless, these will be discussed briefly in this section.

In essence, synchronous discussions might be conducive to high-quality discussions since they can be regarded as being online versions of offline deliberation events, which are designed to establish an “ideal speech situation” (see Habermas 1996). As Noveck (2004, 27) writes: “[...] we wanted to mimic in cyberspace the effect of getting people together in the same place at the same time to confront new ideas. Hence, it [her experiment, authors’ note] had to be a synchronous application”. Online deliberative polls (for example, Fishkin 2009; Luskin, Fishkin, and Iyengar 2006) and some other experiments (cf. Grönlund et al. 2009; Strandberg and Grönlund 2012) have had similar ideas of creating virtual versions of offline deliberations. Although it has not been explicitly written in any of their studies, it is perceivable that these scholars recognise a value *vis-à-vis* discussion quality and deliberation in the interaction between citizens (cf. Smith, John, and Sturgis 2013, 712). After all, reciprocity—that is, facing different views and actively engaging in discussion with citizens who hold these opposing views—is one of the main aspects of an ideal-speech situation (Kies 2010, 40–53). However, it should be acknowledged that synchronous discussions allow less time for contemplation, which could have an adverse effect on discussion quality (for example, Janssen and Kies 2005, 321).

Turning to asynchronous discussions, then, we already mentioned that these could be more beneficial to discussion quality—and to deliberative discussion quality specifically—since they allow time for reflection and contemplation (Janssen and Kies 2005, Smith, John, and Sturgis 2013). Critical-rational discourse, reasoned justification of arguments and reflexivity are all very central elements in a good-quality deliberative discussion (for example, Dahlberg 2001; Stromer-Galley 2007), all of which are significantly aided by having sufficient time for contemplation. In this respect, asynchronous forums are also regarded as being more accessible since people can take part in a discussion at their own convenience (Davies et al. 2009, 283), which might serve to broaden the range of citizens engaging in discussion and increase the likelihood of encountering different views and having to rationally defend one’s own views. This, in turn, could enhance discussion quality regarding both reciprocity and the critical rationality. Nonetheless, the weakness of asynchronous discussions is that they are less likely, *a priori*, to have the same intensity of reciprocity between participants as synchronous discussions (Smith, John, and Sturgis 2013, 710).

### *Potential Impact of Identifiability on Discussion Quality in Online Deliberations*

In a sense, it is understandable that the identifiability of participants—that is, whether their identities are known to each other or not—has not received any attention in offline deliberation studies. After all, face-to-face discussion is hard to conduct in a credible way without participants at least being able to see and hear each other. Indirectly, however, scholars have touched upon the topic in discussions of how status and social position often have a negative influence on citizens’ discussions (for example, Gastil 2000, 359) or

how people may feel socially uncomfortable in the deliberative setting (cf. Mutz 2006). Thus, it is perceivable that being non-identifiable might have positive effects for certain citizens when deliberating (Stromer-Galley 2002, 35–38), although it should be stressed that the social effects of anonymity are nonetheless ambiguous and context dependent (Kahai 2009, 451–452).

Nevertheless, the general relevance of identifiability has clearly increased due to the development of digital communication technology. For instance, whereas anonymity is rare and requires much effort to achieve offline, it is more or less the prevailing routine online (Samuel 2004). Therefore, it is quite surprising that despite a vast amount of research into the quality of online discussions in general (for example, Papacharissi 2004), and also a growing field of research in online deliberation (for example, Esterling, Neblo, and Lazer 2011; Luskin, Fishkin, and Iyengar 2006; Muhlberger 2005; Price and Capella, 2002), very few studies have actually directly considered the effects that identifiability might have in online communication, let alone in online deliberation. Still, some reflections and findings regarding its influence in interpersonal communication, which are of relevance to online deliberation too, do exist.

Firstly, Samuel (2004, 214) remarks that being anonymous shifts the focus of the discussion onto what is being said rather than who is saying it. Thus, the force of the better argument, which is central to deliberative theory (for example, Habermas 1996), would be more likely to prevail. Additionally, being anonymous might free people from psychological barriers which have often held back their engagement in deliberation. Stromer-Galley (2002, 35–38), for instance, found that the feeling of being private while typing on your computer at home, while nonetheless engaging in public political conversations online, is suitable for citizens who might feel uncomfortable in face-to-face settings. Likewise, scholars have recognised that the online setting could reduce the influence of status and social position on citizens' discussions (for example, Dahlberg 2001, 14; Price 2009, 43; Witschge 2004, 116). Thus, people may actually feel more comfortable expressing politically incorrect opinions in an anonymous environment (Tucey 2010). Online anonymity may also lessen the effect of the spiral of silence (Noelle-Neumann 1993) in the sense of alleviating minorities' fears of social isolation, thus giving them greater opportunities to express divergent views (cf. Price 2009, 43). Concerning this, Leshed (2009), for instance, found that the removal of the possibility of anonymity decreased both the number of posts and the presence of dialogue in a corporate online community.

However, poor discussion quality has been a major concern with online discussions, and it has often been suggested that one of the main causes of this is anonymity (for example, Coleman and Moss 2012, 8)—something which has also been empirically demonstrated (for example, Nagar 2011; Santana 2014). Tucey (2010) claims that anonymous discussion is often of a lower quality—leading to flaming, rudeness and less thoughtful contributions to the debate—than discussions where identities are known, because participants may feel that they are not responsible for their words. In a similar vein, psychological research has demonstrated that anonymity may lead to disinhibition and de-individualisation, which can make persons act in ways that violate the social norm (see Wallace 2008). Conversely, a discussion where messages can be linked to individuals may create a sense of responsibility and desire to have a serious discussion. Sundström (2002) points out that a discussant's status does not only have a negative influence on citizens' discussions, *per se*; it sometimes might actually strengthen the argument made. Thus,

even though anonymity erodes the influence of status (for example, Price 2009, 43; Witschge 2004, 116), it is nonetheless unclear whether this actually ensures better discussion quality.

## Research Questions

Thus far, we have touched upon certain criteria of discussion quality in the theoretical backdrop. In order to formulate research questions for the study, however, we will first need to elaborate on deliberative discussion quality. Initially, it should be stressed that scholars are yet to reach consensus on what constitutes a high-quality deliberation and how to measure it (see Davies and Chandler 2011, 9). Indeed, dimensions of deliberative discussion quality are multifaceted and complex (Stromer-Galley 2007; Trénel 2004) and would arguably merit an article of their own in order to thoroughly discuss all the relevant aspects. Within the limited scope of this article, however, we concentrate on commonly used indicators of discussion quality (Graham 2009; Stromer-Galley 2007; Trénel 2004)—that is, the rationality of the discussions, the relevance to the topic of the discussions, the reciprocity of the discussions and the degree of politeness and respect in the discussions—in order to focus on core aspects of discussion quality in deliberation.

As indicated in the discussion contained within the theoretical backdrop, both temporality and identifiability appear to have relevance to discussion quality generally, and deliberative discussion quality specifically. In spite of this, however, very little empirical research has directly and systematically tested the impact of these factors on discussion quality in a deliberative context. As clear-cut linkages between temporality and identifiability were hard to discern, we found it difficult to formulate hypotheses to guide the empirical part of our study. Instead, the following array of open-ended research questions will be addressed empirically:

- RQ1: How do temporality and identifiability affect the rationality of discussions in an online deliberative context?
- RQ2: How do temporality and identifiability affect the relevance of discussions in an online deliberative context?
- RQ3: How do temporality and identifiability affect the reciprocity of discussions in an online deliberative context?
- RQ4: How do temporality and identifiability affect the degree of politeness and respect of discussions in an online deliberative context?

## Method and Data

### *Design*

To test the impact of temporality and identifiability on discussion quality, four different discussion settings in an online deliberation were used as treatment stimuli in an online experiment. Thus, a  $2 \times 2$  factorial design representing two levels of the temporality factor (i.e. asynchronous vs. synchronous) and two levels of the identifiability factor (i.e. anonymous vs. known identities) was applied in the experiment (Figure 1).

The impact on discussion quality was examined for both main effects (i.e. the independent effect of each factor) and interaction effects (i.e. effects occurring when certain levels of both factors combine). The experiment used the classic post-test-only design

	Asynchronous discussion	Synchronous discussion
Anonymous discussants	1 ( <i>n</i> = 16)	2 ( <i>n</i> = 13)
Known identity	3 ( <i>n</i> = 12)	4 ( <i>n</i> = 9)

**FIGURE 1**  
Experimental design: 2 × 2 factorial design

with between-subjects comparisons (Campbell and Stanley 1963; Gribbons and Herman 1997). Accordingly, an initial pool of participants was firstly recruited and thereafter randomly assigned to experimental treatments. Participants are thus regarded as equivalent to each other before being exposed to the treatments, and any differences in discussion quality observed between the treatment groups are logically contingent on the different contextual features of each treatment. However, due to the low number in our experiment, we double-checked that the randomisation had truly succeeded prior to the actual experiment beginning to ensure that the post-test-only design would remain valid (see Gribbons and Herman 1997). None of seven background variables—age, gender, occupation, education, the degree of urbanity of area of residence, interest in politics and the extent to which the participants usually engage in political discussions—displayed any statistically significant between-group differences (chi-squares test of distributions, lowest *p* value found in the tests was 0.19).

### Participants

The final number of volunteers taking part in the discussions was 50. The random allocation of participants to treatment groups yielded on average 13 participants per group<sup>1</sup>. This resulted in an approximated *a priori* effect size for the experiment of 0.50 (Cohen’s *d*) while maintaining a high statistical power of 0.89 on the 0.90 significance level<sup>2</sup> (cf. VanVoorhis and Morgan 2007, 48).

The participants were aged between 17 and 68 (mean age was 28); 53 per cent were female while 47 per cent were male. As recommended by Druckman and Kam (2011, 54), we recruited a dual sample of graduate students and non-students. Although a representative sample is often preferable in deliberation, for instance to reduce the risk of enclave discussions where only like-minded people discuss with each other (Fishkin 2009, 33), the use of students was necessary since the participation in this experiment required extensive online skills in the use of webcams and headset equipment in the synchronous/known identity group; skills which are common among young people (cf. Kies 2010, 67). In the final sample, 59 per cent of the participants were students. As remarked by Druckman and Kam (2011, 41), the use of students in experiments is often heavily criticised but the criticism is misguided (see Druckman and Kam [2011] for elaboration and empirical evidence in support of this claim). Nonetheless, it has to be acknowledged that there is a risk of so-called crystallisation in the sense that graduate students may be relatively strong in skills required for deliberating, which could mean that the experiment fails to pick up between-group

differences in discussion quality since there was too little variation within groups in a crystallising factor; that is, deliberative capacity (Druckman and Kam 2011, 45–46). In other words, the characteristics of the participants may “triumph” over the discussion setting and blur the effects of contextual factors on discussion quality. This will be taken into account in the presentation of findings from the experiment.

Participants volunteered for the experiment on the study’s website. At this stage, basic demographic information on the participants was also collected. All participants were promised a gift voucher worth €50 upon finishing the entire experiment, and to be entered into a lottery for a weekend boat cruise. Of course, this non-probability-based recruitment method is vulnerable to volunteering, which could decrease the representativeness of the sample. However, it is important to note that the purpose of this experiment is not to achieve generalisation about how citizens discuss online, in which case the demographic representativeness of the sample would have been crucial. Rather, the focus here is explicitly on studying the effects of the variation in discursive settings; what happens to the discussion quality if we vary the discursive setting concerning two factors? Additionally, the experiment’s population could even be seen as a rather critical test (cf. Druckman and Kam 2011, 53); if between-treatment differences in discussion quality are found even though the characteristics of the population make these unlikely to occur, it is to be logically expected that the same manipulation would also yield an effect on a more representative population where, for instance, the risks of crystallisation or enclave deliberation are smaller.

### *Stimulus*

The stimulus consisted of taking part in online discussions which were set up according to the typology depicted in Figure 1. The discussion topic concerned whether same-sex couples should have the right to adopt children—that is, a rather controversial issue in Finland (see Strandberg *forthcoming*). Each participant only participated in one type of discussion. All discussions were held on the study’s website. Similarly to some earlier online deliberation studies (for example, Noveck 2004; Luskin, Fishkin, and Iyengar 2006; Muhlberger 2005; Price and Capella 2002), the synchronous discussions (Cells 2 and 4 in Figure 1) were modelled after offline mini-publics. Thus, these were live discussion sessions supervised by a trained facilitator, who was tasked with ensuring that the flow of discussion was good and that everyone participated, and with providing technical assistance. The anonymous synchronous discussion (Cell 2 in Figure 1) was text-only using pseudonyms (i.e. Participant 1, 2, 3, etc.) for each participant. The known-identity synchronous discussion (Cell 4) used webcams and microphones and each participant’s identity (i.e. their full name) was known to the other participants. The asynchronous discussion treatment groups (Cells 1 and 3 in Figure 1) took part in a three-week text-based discussion on a discussion board. These participants received a topic of discussion but no facilitator needed to supervise the discussion. The anonymous asynchronous discussants (Cell 1) used pseudonyms (e.g. Participant 1, 2, 3, etc.) whereas the asynchronous discussions with known identities (Cell 3) used participants’ real names as well as photographs.

It has to be stressed that the ways in which the discussions were set up—albeit due to technical and budgetary limitations—introduced additional factors (i.e. text vs. audio/video and facilitator vs. no facilitator), which varied between the groups. Since these aspects may

potentially cause variations in discussion quality and thus damage the internal validity of the experiment, they will be taken into account when interpreting the findings.

### *Dependent Variables*

Discussion quality was measured by content analysing (quantitative content analysis) all speeches and posts in the discussions ( $n = 707$ ). The two asynchronous discussion groups had 38 posts (2.4 posts per participant) and 36 posts (3 posts/participant) respectively, all of which were very long (average 916 and 1093 characters respectively). The two synchronous groups had a much higher number of speeches/posts—399 (31 posts/participant) and 234 respectively (26 posts/participant)—which were much shorter in length (averaging 85 and 335 characters respectively). As mentioned in the section on research questions, the quality of discussion was measured using four dimensions: rationality; relevance; reciprocity; and politeness and respect.

Rationality concerns the extent to which debaters make efforts to logically and rationally defend and elaborate their claims. Relevance basically means staying on topic (see Jensen 2003; Stromer-Galley 2007; Graham 2009). Reciprocity can be likened to Barber's strong democratic talk: "[...] always involves listening as well as speaking, feeling as well as thinking, and acting as well as reflecting" (1984, 178). Politeness and respect, finally, is both a precondition for the democratic quality of conversation and at the same time signs of quality in such conversations (Papacharissi 2004). In our final coding scheme, we operationalised these criteria through the use of several different indicators adapted from the literature (cf. Graham 2009; Steenberger et al. 2003; Strandberg and Berg 2013). Each of the main criteria of discussion quality as well as their operationalisations are summarised in Table 1, in which the inter-coder reliability (Krippendorff's alpha) is also provided for each variable.

**TABLE 1**  
Criteria, variables and intercoder reliability

Criteria	Variable	Krippendorff's alpha
Rationality	Opinion justification level	0.709
	General justification level	0.768
	Issue opinion quality	0.714
	General opinion quality	0.828
	External validation	1
	Internal validation	1
	Argumentation	1
Relevance	Relevance	0.894
Reciprocity	Agreement	0.808
	Disagreement	1
	Conviction	1
	Progress	0.808
	Radicalisation	1
Politeness and respect	Tone	0.761
	Incivility	1
	Impoliteness	1

Note: Two coders coded each variable for a randomly-selected subsample of 10 posts. If a Krippendorff's alpha value of 0.7 was not met for a variable, an additional coding was performed using a new subsample and revised instructions. The total intercoder agreement for the range of variables was 84 per cent.

### *Procedure*

The experiment took place in November 2013 on an online platform for experimental research. By using an online platform we hoped to increase the experiment's external validity since the experiment is to be regarded as an online lab-in-the-field experiment (Morton and Williams 2010, 296–297). The setting thus retains some of the benefits of a controlled laboratory experiment, such as a relatively high internal validity, and some of the higher external relevance—beyond the laboratory—of a field experiment.

The experiment included three stages; a recruitment stage; the random allocation to treatment groups; and an experimental stage. The true purpose of the experiment was initially unknown to the participants in order to avoid goal-directed behaviour. Thus, the participants did not know about the different treatments and were only told that we were interested in studying online discussion. When all the participants had taken part in their respective discussions they were debriefed on the actual purpose of the experiment according to good ethical practice in experimental research.

### **Findings**

Although there were a total of 707 posts to be analysed, initial data analyses showed that the data were not normally distributed (Kolmogorov–Smirnov test of normality). Consequently, in order to analyse the main effects and interactions in the full  $2 \times 2$  factorial design, all scores on the indicators of discussion quality were transformed into ranks before analysis by means of the adjusted rank-transform method. The benefit therein is that parametric methods can be used on non-normally distributed data since rank transformation is carried out *ex ante facto* instead of *post facto*, as is the case in non-parametric methods (see Leys and Schumann 2010). Thus, the analysis is performed through multivariate analysis of variance concerning discussion quality as a whole, including univariate analyses of variance for each dependent variable separately. Statistical significance is indicated at the 0.10 level to ensure a very high statistical power of 0.89 in the experiment. The findings are presented for each factor separately followed by presentation of the interaction effects.

#### *Main Effects on Discussion Quality of Factor 1, Temporality, in the $2 \times 2$ Factorial Design*

We start by presenting the findings concerning the main effects on discussion quality of the first factor, temporality, in the  $2 \times 2$  factorial design (Table 2).

Evidently, temporality had a significant main effect ( $F(8,696) = 22.632, p = 0.000, \eta p^2 = 0.206$ , observed power = 1.00) on discussion quality as a whole, as indicated by the results of the multivariate analysis presented in Table 2. Peering further into its effects on each of the indicators of discussion quality—by looking at the results from the univariate analyses—the asynchronous level of the temporality factor has the most positive main effects on discussion quality. Thus, for all of the five indicators which we used to gauge rationality, there were positive main effects for asynchronous discussions. The effect sizes for these ranged between 0.009 and 0.095 and observed power between 0.790 and 1.00. There was no statistically significant main effect for temporality concerning the relevance indicator but asynchronous discussions had significant main effects on both

**TABLE 2**  
Main effects of the temporality factor on discussion quality

	<i>F</i>	<i>df</i>	<i>p</i>	$\eta^2$	Observed power	Group with highest mean
<b>Multivariate analysis</b>	22.632	8, 696	0.000	0.206	1.00	
<b>Univariate analyses</b>						
<i>Rationality</i>						
Opinion justification	20.344	1, 703	0.000	0.028	0.998	Asynchronous
General justification	73.602	1, 703	0.000	0.095	1.00	Asynchronous
Issue opinion quality	6.028	1, 703	0.014	0.009	0.790	Asynchronous
General opinion quality	8.176	1, 703	0.004	0.011	0.887	Asynchronous
Validations and argumentation	71.999	1, 703	0.000	0.093	1.00	Asynchronous
<i>Relevance</i>	n.s.	n.s.	n.s.	n.s.	n.s.	
<i>Reciprocity</i>	5.425	1, 703	0.020	0.008	0.752	Asynchronous
<i>Politeness and respect</i>	30.486	1, 703	0.000	0.042	1.00	Asynchronous

reciprocity ( $F(1,703) = 5.425, p = 0.020, \eta^2 = 0.008$ , observed power = 0.752) and politeness and respect ( $F(1,703) = 30.486, p = 0.000, \eta^2 = 0.042$ , observed power = 1.00).

*Main Effects on Discussion Quality of Factor 2, Identifiability, in the 2 × 2 Factorial Design*

We now continue by examining the main effects on discussion quality of the second factor—identifiability—in the 2 × 2 factorial design (Table 3).

In comparison with the findings concerning the temporality factor, the main effects of the identifiability factor are very few and weak. Hence, although identifiability has a weak main effect of discussion quality as a whole ( $F(8,696) = 3.607, p = 0.000, \eta^2 = 0.040$ , observed power = 0.994), only one significant main effect was found for any of the individual indicators of discussion quality. That effect was concerning general justifications where the known identity level brought about higher quality ( $F(1,703) = 2.852, p = 0.092, \eta^2 = 0.004$ , observed power = 0.517). However, since the alpha value (the *p* value) was rather

**TABLE 3**  
Main effects of the identifiability factor on discussion quality

	<i>F</i>	<i>df</i>	<i>p</i>	$\eta^2$	Observed power	Group with highest mean
<b>Multivariate analysis</b>	3.607	8, 696	0.000	0.040	0.994	
<b>Univariate analyses</b>						
<i>Rationality</i>						
Opinion justification	n.s.	n.s.	n.s.	n.s.	n.s.	
General justification	2.852	1, 703	0.092	0.004	0.517	Known identity
Issue opinion quality	n.s.	n.s.	n.s.	n.s.	n.s.	
General opinion quality	n.s.	n.s.	n.s.	n.s.	n.s.	
Validations and argumentation	n.s.	n.s.	n.s.	n.s.	n.s.	
<i>Relevance</i>	n.s.	n.s.	n.s.	n.s.	n.s.	
<i>Reciprocity</i>	n.s.	n.s.	n.s.	n.s.	n.s.	
<i>Politeness and respect</i>	n.s.	n.s.	n.s.	n.s.	n.s.	

high, the effect size was very low and the observed power was equally very low, the significant effect must be regarded as very weak.

*Interaction Effects on Discussion Quality of Factor 1, Temporality, and Factor 2, Identifiability, in the 2 × 2 Factorial Design*

Before the concluding discussion, we look at the interactions between the two factors in the 2 × 2 factorial design (Table 4).

**TABLE 4**  
Interaction effects on discussion quality of the temporality and identifiability factors

	<i>F</i>	<i>df</i>	<i>p</i>	$\eta^2$	Observed power	Group with highest mean
<b>Multivariate analysis</b>	4.203	8, 696	0.000	0.046	0.998	
<b>Univariate analyses</b>						
<i>Rationality</i>			0			
Opinion justification	10.298	1, 703	0.001	0.014	0.941	Asynchronous × anonymous
General justification	n.s.	n.s.	n.s.	n.s.	n.s.	
Issue opinion quality	n.s.	n.s.	n.s.	n.s.	n.s.	
General opinion quality	n.s.	n.s.	n.s.	n.s.	n.s.	
Validations and argumentation	n.s.	n.s.	n.s.	n.s.	n.s.	
<i>Relevance</i>	n.s.	n.s.	n.s.	n.s.	n.s.	
<i>Reciprocity</i>	n.s.	n.s.	n.s.	n.s.	n.s.	
<i>Politeness and respect</i>	n.s.	n.s.	n.s.	n.s.	n.s.	

The findings concerning interaction effects were very similar to those concerning the identifiability factor. Thus, there was a significant weak interaction effect regarding overall discussion quality found in the multivariate analysis ( $F(8,696) = 4.203$ ,  $p = 0.000$ ,  $\eta^2 = 0.046$ , observed power = 0.998), and only one significant interaction effect concerning the specific indicators of discussion quality. The latter of these was regarding opinion justifications where the asynchronous discussion with anonymous discussants displayed the highest quality of all treatment groups ( $F(1,703) = 10.298$ ,  $p = 0.001$ ,  $\eta^2 = 0.014$ , observed power = 0.941). It should be noted, however, that the effect size of this interaction was very low.

## Conclusions

In this article, we asked a question concerning how discussion quality in an online deliberation is affected if we stray from the ideal-speech situation by manipulating the contextual features of the discussion. Although we have only tested two factors in this study—that is, temporality and identifiability—certain conclusions are nonetheless evident in light of our findings.

Firstly, the template for deliberation used in offline mini-publics (for example, Fishkin, Luskin, and Jowell 2000)—that is, live events with known participants discussing relevant

topics over a few hours—is often seen as resembling an ideal-speech situation. However, our findings would tentatively suggest that straying from such a setting in an online deliberation is not necessarily negative for discussion quality. Thus, neither the synchronous level of the temporality factor nor the known identity level of the identifiability factor was especially influential on discussion quality in our experiment. In fact, both of these combined only had one weak statistically significant positive main effect on quality (known identity concerning general justifications). However, more firm conclusions emerge when moving beyond the rather vaguely formulated general question towards the specific research questions regarding the influence on discussion quality of our two examined factors.

Hence, our most noteworthy finding is that the asynchronous level of the temporality factor had significant main effects for all but one indicator of discussion quality, and was also part of the only significant interaction effect found. Thus, similar to what some other studies of online deliberation have suggested (for example, Janssen and Kies 2005; Smith, John, and Sturgis 2013), having sufficient time for reflection and contemplation—even when discussing anonymously—did indeed turn out to be decisive for discussion quality upon systematic empirical investigation. Regarding this, it is also noteworthy that this held true even for the reciprocity indicator—one which we expected to be aided by synchronous temporality and not asynchronous temporality. Likewise, it is noteworthy that the asynchronous discussions in our experiment would have been less likely *a priori* to have good discussion quality since they did not have any facilitation, whereas the synchronous discussions had. We would draw the conclusion from these findings that citizens' discussions online are affected by different contextual features than discussions offline; for instance, asynchronous deliberation offline would be very unnatural and hardly lead to good discussion quality, whereas it is part of the communicative fabric of the online realm and therefore a natural fit for deliberation too. From a methodological perspective, also, asynchronous discussions are technically easier to conduct online than synchronous discussions. Thus, albeit a natural starting point of an emerging research field, the quite common practice of “translating” offline deliberation events into online counterparts—which has also been investigated by one of the authors of this article (Grönlund et al. 2009; Strandberg & Grönlund 2012)—might not be the best way to realise online deliberation when seeking the best deliberative discussion quality.

Although we generally did not find many effects on discussion quality for the identifiability factor, we would argue that this nonetheless tentatively indicates that adjusting the communicative context of online deliberations to the core features of the communication medium might be more important for discussion quality than replicating offline deliberations *per se*. Thus, anonymity has often been regarded more or less as the *modus operandi* of the Internet (Samuel 2004), which has had a mostly comparable influence on discussion quality when compared with the known identity of discussants—that is, both had almost no impact—in our online deliberation. Of course, the findings of this study would have to be replicated in future experiments, which employ larger and more representative experimental populations and are conducted in a broad range of contexts, in order to know how generally applicable this conclusion is.

On the note of generalisability, some limitations of the study need attention. Thus, while we argued that the online lab-in-the-field experiment would be beneficial to both

internal and external validity (Morton and Williams 2010), both of these had *de facto* shortcomings. Internal validity, firstly, was not perfect due to extraneous factors—that is, audio/video versus text only and facilitation versus no facilitation—being built into the experiment owing to technical limitations. However, as we have discussed, this actually served to strengthen the main finding regarding asynchronous discussions. External validity, secondly, can also be questioned since the population of the experiment was not representative of the general Finnish population, notwithstanding that it was perhaps a critical test. Thus, the experiment would need to be replicated with a larger and more representative population to increase the external validity of the findings. A third limitation concerns the validity of our measurements of discussion quality. Although we have focused on what could be regarded as the common denominators within the research field (for example, Graham 2009; Kies 2010, 40–53), there are many alternative indicators yet to be studied (cf. Stromer-Galley 2007; Trénel 2004).

Wrapping up the study, it is important to note that one experiment only goes so far in furthering our knowledge. Thus, as stated, more experimental studies need to be conducted, using both more representative samples and other countries as context, in order to ensure that our findings are not only robust but also gain new knowledge on other factors relevant to discussion quality in online deliberations. For instance, the influence on discussion quality of the two extraneous factors which appeared in this study—audio/video versus text and facilitation versus no facilitation—as well as a range of other factors still needs more systematic empirical attention. Other studies, for instance, have pointed to the relevance of transparency on discussion quality (Stasavage 2007)—that is, whether the discussion is public or private—the size of the actual discussion groups (Undem 2001) as well as whether the number of contributions per day is limited, in order to prevent domination by only a few participants (Coleman and Moss 2012, 8). Additionally, although we initially argued that this study was one of only a few actually measuring discussion quality in deliberations, and not only so-called democratic outcomes thereof, there is still a need for studies which connect the relevant factors in the deliberative design, the actual discussion quality and the democratic outcomes.

## DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

## FUNDING

This work was funded by the Academy of Finland. Project nr. 137813.

## NOTES

1. The number of participants was not equal in all cells due to some participants dropping out at the very last moment before the discussion.
2. The precise figures were calculated with the aid of an online calculator: <http://www.stat.ubc.ca/~rollin/stats/ssize/n2.html>.

## REFERENCES

- Albrecht, Stephen. 2006. "Whose Voice is Heard in Online Deliberation?: A Study of Participation and Representation in Political Debates on the Internet." *Information, Community and Society* 9 (1): 62–82.
- Bächtiger, André. 2011. "Contestatory Deliberation." Paper presented at the Epistemic Democracy Conference, Yale University, New Haven, CT, October 22.
- Barber, Benjamin. 1984. *Strong Democracy: Participatory Politics for a New Age*. Los Angeles: University of California Press.
- Campbell, Donald, and Julian Stanley. 1963. *Experimental and Quasi-experimental Designs for Research*. Chicago, IL: Rand McNally.
- Coleman, Stephen, and Giles Moss. 2012. "Under Construction: The Field of Online Deliberation Research." *Journal of Information Technology & Politics* 9 (1): 1–15.
- Dahl, Robert. 1989. *Demokratin och Dess Antagonister* [Democracy and its critics]. New Haven, CT: Yale University Press.
- Dahlberg, Lincoln. 2001. The Internet and Democratic Discourse—Exploring the Prospects of Online Deliberative Forums Extending the Public Sphere. *Information, Communication & Society* 4 (4): 615–633.
- Davies, Todd, and Seeta Peña Gangadharan. 2009. *Online Deliberation—Design, Research, and Practice*. San Francisco, CA: CSLI Publications.
- Davies, Todd, and Reid Chandler. 2011. "Online Deliberation Design: Choices, Criteria, and Evidence." In *Democracy in Motion. Evaluating the Practice and Impact of Deliberative Civic Engagement*, edited by T. Nabatchi, J. Gastil, M. Leighninger, and G.M. Weiksner, 103–131. Oxford: Oxford University Press.
- Davies, Todd, Brendan O'Connor, Alex Cochran, Jonathan Effrat, Andrew Parker, Benjamin Newman, and Aaron Tam. 2009. "An Online Environment for Democratic Deliberation: Motivations, Principles, and Design." In *Online Deliberation. Design, Research, and Practice*, edited by T. Davies and S.P. Gangadharan, 275–292. Stanford, CA: CSLI Publications.
- Druckman, James, and Cindy Kam. 2011. "Students as Experimental Participants. A Defence of the 'Narrow Data Base'." In *Cambridge Handbook of Experimental Political Science*, edited by J. Druckman, D. Green, J. Kuklinski and A. Lupia, 41–57. Cambridge: Cambridge University Press.
- Esterling Kevin, Michael Neblo, and David Lazer. 2011. "Means, Motive, and Opportunity in Becoming Informed About Politics. A Deliberative Field Experiment With Members of Congress and Their Constituents." *Public Opinion Quarterly* 75 (3): 483–503.
- Fishkin, James. 2009. "Virtual Public Consultation: Prospect for Internet Deliberative Democracy." In *Online Deliberation. Design, Research, and Practice*, edited by T. Davies and S.P. Gangadharan, 23–35. Stanford, CA: CSLI Publications.
- Fishkin, James, Robert Luskin, and Roger Jowell. 2000. "Deliberative Polling and Public Consultation." *Parliamentary Affairs* 53 (4): 657–666.
- Gastil, John. 2000. "Is Face-to-face Deliberation a Luxury or a Necessity." *Political Communication* 17 (4): 357–361.
- Graham, Todd. 2009. "What's Wife Swap Got to do With it?" Doctoral Thesis, University of Amsterdam.
- Gribbons, Barry, and Joan Herman. 1997. "True and Quasi-experimental Designs." *Practical Assessment, Research & Evaluation* 5, 14.

- Grönlund, Kimmo, Kim Strandberg, and Staffan Himmelroos. 2009. "The Challenge of Deliberative Democracy Online – A Comparison of Face-to-face and Virtual Experiments in Citizen Deliberation." *Information Polity* 14 (3): 187–201.
- Grönlund, Kimmo, Maija Setälä, and Kaisa Herne. 2010. "Deliberation and Civic Virtue Lessons from a Citizen Deliberation Experiment." *European Political Science Review* 2 (1): 95–117.
- Habermas, Jürgen. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge: Polity Press.
- Janssen, Davy, and Raphaël Kies. 2005. "Online Forums and Deliberative Democracy." *Acta Politica* 40 (3): 317–335.
- Jensen, Jakob. 2003. "Public Spheres on the Internet: Anarchic or Government-Sponsored—A Comparison." *Scandinavian Political Studies* 26 (4): 349–374.
- Kahai, Surinder. 2009. "Anonymity and Counter-normative Arguments in Computer-Mediated Discussions." *Group & Organization Management* 34 (4): 449–478.
- Kies, Raphaël. 2010. *Promises and Limits of Web-deliberation*. New York: Palgrave Macmillan.
- Leshed, Gilly. 2009. "Silencing the Clatter: Removing Anonymity from a Corporate Online Community." In *Online Deliberation. Design, Research, and Practice*, edited by T. Davies and S.P. Gangadharan, 243–251. Stanford, CA: CSLI Publications.
- Leys, Christophe, and Sandy Schumann. 2010. "A Nonparametric Method to Analyse Interactions: The Adjusted Rank Transform Test." *Journal of Experimental Social Psychology* 46 (4): 684–688.
- Luskin, Robert, James Fishkin, and Shanto Iyengar. 2006. *Considered Opinions on U.S. Foreign Policy: Evidence From Online and Face-to-face Deliberative Polling*. California: The Center for Deliberative Democracy.
- Manosevitch, Edith. 2014. "The Design of Online Deliberation: Implications for Practice, Theory and Democratic Citizenship." *Journal of Public Deliberation* 10 (1), article 9. URL: <http://www.publicdeliberation.net/cgi/viewcontent.cgi?article=1325&context=jpd>
- Morton, Rebecca, and Kenneth Williams. 2010. *Experimental Political Science and the Study of Causality. From Nature to the Lab*. Cambridge: Cambridge University Press.
- Muhlberger, Peter. 2005. "The Virtual Agora Project: A Research Design for Studying Democratic Deliberation." *Journal of Public Deliberation* 1 (1): article 5. <http://www.publicdeliberation.net/jpd/vol1/iss1/art5/>
- Mutz, Diana. 2006. *Hearing the Other Side: Deliberative Versus Participatory Democracy*. New York: Cambridge University Press.
- Nagar, Na'ama. 2011. "The Loud Public: The Case of User Comments in Online News Media." Doctoral thesis, State University of New York.
- Noelle-Neumann, Elisabeth. 1993. *The Spiral of Silence: Public Opinion—Our Social Skin*. Chicago: University of Chicago Press.
- Noveck, Beck. 2004. "Unchat: Democratic Solution for a Wired World." In *Democracy Online: The Prospects for Political Renewal through the Internet*, edited by P. Shane, 21–34. New York: Routledge.
- Papacharissi, Zizi. 2004. "Democracy Online: Civility, Politeness, and the Democratic Potential of Online Political Discussion Groups." *New Media & Society* 6 (2): 259–283.
- Price, Vincent. 2009. "Citizens Deliberating Online: Theory and Some Evidence." In *Online Deliberation. Design, Research, and Practice*, edited by T. Davies and S. P. Gangadharan, 37–58. Stanford, CA: CSLI Publications.
- Price, Vincent, and Jamie Capella. 2002. "Online Deliberation and Its Influence: The Electronic Dialogue Project in Campaign 2000." *IT & Society* 1 (1): 303–329.

- Rhee, June, and Eun-Mee Kim. 2009. "Deliberation on the Net: Lessons From a Field Experiment." In *Online Deliberation. Design, Research, and Practice*, edited by T. Davies and S. P. Gangadharan, 223–232. Stanford, CA: CSLI Publications.
- Ryfe, David. 2005. "Does Deliberative Democracy Work?" *Annual Review of Political Science* 8: 49–71.
- Samuel, Alexandra. 2004. "Hacktivism and the Future of Political Participation.2 Doctoral thesis, Harvard University, Cambridge, MA.
- Santana, Arthur. 2014. "Virtuous or Vitriolic: The Effect of Anonymity on Civility in Online Newspaper Reader Comment Boards." *Journalism Practice* 8 (1): 18–33.
- Smith, Graham, Peter John, and Patrick Sturgis. 2013. "Taking Political Engagement Online: an Experimental Analysis of Asynchronous Discussion Forums." *Political Studies* 61 (4): 709–730.
- Stasavage, David. 2007. "Polarization and Publicity: Rethinking the Benefits of Deliberative Democracy." *The Journal of Politics* 69 (1): 59–72.
- Steenberger, Marco, André Bächtiger, Markus Spröndli, and Jürg Steiner. 2003. "Measuring Political Deliberation: A Discourse Quality Index." *Comparative European Politics* 2003 (1): 21–48.
- Strandberg, Kim, and Kimmo Grönlund. 2012. "Online Deliberation and Its Outcome – Evidence from the Virtual Polity Deliberative Experiment." *Journal of Information Technology and Politics* 9 (2): 167–184.
- Strandberg, Kim, and Janne Berg. 2013. "Online Readers' Comments – Democratic Conversation Platforms or Virtual Soapboxes?" *Comunicao e Sociedade - Communication & Society* 23: 132–152.
- Strandberg, Kim, and Kimmo Grönlund. 2013. "Online Deliberation – Theory and Practice in Virtual Mini-publics." In *Deliberative Mini-Publics – Practices, Promises, Pitfalls*, edited by K. Grönlund, A. Bächtiger, and M. Setälä. Colchester: ECPR Press.
- Strandberg, Kim. forthcoming. "Designing for Democracy? – An Experimental Study Comparing the Outcomes of Citizen Discussions in Online Forums to Those of a Designed Online Citizen Deliberation Venue." *European Political Science Review*.
- Stromer-Galley, Jennifer. 2002. "New Voices in the Public Sphere: A Comparative Analysis of Interpersonal and Online Political Talk." *Javnost—The Public* 9 (2): 23–42.
- Stromer-Galley, Jennifer. 2007. "Measuring Deliberation's Content: A Coding Scheme." *Journal of Public Deliberation* 3 (1): Article 12. <http://www.publicdeliberation.net/cgi/viewcontent.cgi?article=1049&context=jpd>
- Sundström, Mikael. 2002. "Demokratiska Avatarer [Democratic Avatars]." *Statsvetenskaplig Tidskrift* 105 (3): 214–217.
- Towne, Ben, and James Herbsleb. 2012. "Design Considerations for Online Deliberation Systems." *Journal of Information Technology & Politics* 9 (1): 97–115.
- Trénel, Matthias. 2004. "Measuring the Deliberativeness of Online Discussions." Coding Scheme 2.4 Report, Social Science Research Centrex, Berlin.
- Tucey, Cindy. 2010. "Online vs. Face-to-Face Deliberation on the Global Warming and Stem Cell Issues." Paper presented at the Western Political Science Association 2010 Annual Meeting, March 31 – April 3. San Francisco, CA.
- Udem, Tresa. 2001. "Factors Affecting Discussion Quality: The Effects of Group Size, Gender, and Political Heterogeneity in Online Discussion Groups." Unpublished Master's thesis, University of Pennsylvania, Philadelphia.

- VanVoorhis, Carmen, and Betsy Morgan. 2007. "Understanding Power and Rules of Thumb for Determining Sample Sizes." *Tutorials in Quantitative Methods for Psychology* 3 (2): 43–50.
- Wallace, Kathleen. 2008. "Online Anonymity." In *The Handbook of Information and Computer Ethics*, edited by K. Himma K and H. Tavani, 165–190. Hoboken, NJ: Wiley & Sons.
- Witschge, Tamara. 2004. "Online Deliberation: Possibilities of the Internet for Deliberative Democracy." In *Democracy Online: The Prospects for Political Renewal through the Internet*, edited by P. Shane, 109–122. New York: Routledge.

**Kim Strandberg** (corresponding author) is Associate Professor at the social science research institute at the Faculty of Social Sciences, Business and Economics, Åbo Akademi University, Vasa, Finland. Email: [kistrand@abo.fi](mailto:kistrand@abo.fi)

**Janne Berg** is a PhD student at the social science research institute at the Faculty of Social Sciences, Business and Economics, Åbo Akademi University, Vasa, Finland. Email: [janne.berg@abo.fi](mailto:janne.berg@abo.fi)