A Model for the Analysis of Online Citizen Deliberation: Barcelona Case Study

ROSA BORGE BRAVO[1]
JOAN BALCELLS
ALBERT PADRÓ-SOLANET
Universitat Oberta de Catalunya, Spain

Are participatory platforms facilitating public deliberation? To answer this question, we focus on the most commented citizens' proposal discussed on the Barcelona government's platform Decidim (i.e., the granting of new licenses for tourist apartments). Our goal is twofold. First, we evaluate via content analysis the deliberative quality of this conversation through a carefully selected system of indicators following the classical literature on deliberation. Second, we examine how deliberative quality criteria evolve through interaction, by introducing the dimension of depth, inspired on social computing research. The findings show that the relation between deliberative quality and depth of conversation is mostly curvilinear. The level of justification decreases as conversations go deeper, whereas the levels of reciprocity and incivility become more important over time before decreasing at a later stage. Overall, we conclude that online citizen deliberation can spontaneously emerge, but additional institutional conditions are required to make it last.

Keywords: online deliberation, online conversations, deliberative criteria, online participation, participatory platforms, Barcelona, tourism

Following wins by new left-wing parties in Spain's 2015 local elections, new participatory platforms were launched to enhance citizen participation and debate on local public policies. Platforms rolled out in Barcelona and Madrid have been adopted by other Catalan, Spanish, and European cities and have become an example for the public sector.[2] The modular design of these platforms allows local governments to open

Rosa Borge Bravo: rborge@uoc.edu
Joan Balcells: jbalcells@uoc.edu
Albert Padró-Solanet: apadro_solanet@uoc.edu
Date submitted: 2018–10–30

up spaces for deliberation with themed nesting of comments, similar to that of social media. But are these platforms really facilitating public deliberation?

We answer this question from a twofold perspective. Firstly, following the empirical literature on online deliberation, we aim to develop a system of indicators that are able to measure the deliberative capacity of conversations in participatory platforms. This study is a continuity of previous studies on political parties' online fora (Borge & Santamarina, 2016). Thus, we aim to identify which deliberative standards can be meet in specific online media systems and which conditions hamper their development (Bächtiger, Niemeyer, Neblo, Steenbergen, & Steiner, 2010, pp. 45, 59). Secondly, we explore how the different dimensions of deliberative quality vary over time, as conversation trees progressively develop through users' interaction. In the classical literature on deliberation, the depth of the conversations has not been explicitly studied, but the duration of the dialogue affects the unfolding of the deliberative criteria.

Thereby, our approach crosses between two strands of literature. On the one hand, the literature on deliberative quality, focused on the comparison, via content analysis, between deliberative practices and the normative standards of good deliberation; and, on the other hand, the literature on online conversation structures, which relies on Big Data and computational models to describe patterns of interaction.

Our analysis is focused on Barcelona's digital platform Decidim, an open-source platform especially designed to provide citizen input for governments. This platform was first applied to Barcelona's Strategic City Planning, where citizens could send policy proposals to the local government. The article uses as a case analysis the most commented issue in the Strategic City Planning— the granting of new licenses for tourist apartments. This debate tackles a controversial issue, which has arisen not only in Barcelona but also in many other cities affected by global mass tourism. Methodologically, the chosen debate represents a good case analysis, as it is both small enough for content analysis through human coding and large enough to meaningfully analyze its structure through computational techniques.

The Analysis of the Online Deliberation and Its Empirical Criteria

Online deliberation has been analyzed from two perspectives: the most classical one, which focuses on how concrete online conversations fulfil the normative criteria of deliberation (Borge & Santamarina, 2016; Friess & Eilders, 2014; Graham & Witschge, 2006; Kies, 2010; Klinger & Russmann, 2015; Steiner, 2012; Stromer-Galley, 2007), and the most recent one, based on Big Data, which focuses on the structure and development of interactions along multiple conversations (Aragón et al., 2017; Gómez, Kaltenbrunner, & López, 2008; Gonzalez-Bailon, Kaltenbrunner, & Blanchs, 2010). We think that both perspectives are useful and have theoretical and methodological strengths that can be combined to achieve a better understanding of online deliberation. The first perspective follows well-founded theoretical models of communication and democracy and usually analyzes in detail the content of the conversations—but in one moment in time, limited to few cases, and in a noncumulative manner (Gonzalez-Bailon et al., 2010, p. 2). The second perspective derives from network analysis and computer science and examines the structural

features of the conversations, such as width and depth[3] (Aragón et al., 2017; Gonzalez-Bailon et al., 2010) and the network of interactions among participants (Gómez et al., 2008). However, this later perspective applies preconfigured models and algorithms that can be ill-adapted to political conversations, as they focus on the skeleton of the conversation, frequently ignoring the content inside. Interactions' depth and width are only structural measures of the conversation. It is necessary to examine the content of the comments or posts to assess the deliberative quality of a conversation. Our research seeks to bridge the gap between these two perspectives.

Following the first perspective, to assess the deliberative quality of the conversation, we apply widely acknowledged criteria from the classical literature on deliberation in general and online deliberation in particular. In addition, we include in our theoretical framework and analysis some of the structural measures (depth of the conversational threads) and standpoints (evolution of the conversation, interactions' networks) from the second perspective.

Within the first perspective, most of the authors pinpoint that three different levels should be considered to evaluate the deliberation in online platforms (Dahlgren, 2005; Friess & Eilders, 2014, p. 1; Kies, 2010): (1) the institutional and technical dimension of the platform and the space for conversation, (2) the interactive or communicative traits of the conversations, and (3) the outcomes of the online deliberative process. Our analysis is mainly focused on the quality of the communication—that is, the second dimension— though the other two dimensions are also considered because of their relevance for deliberative quality.

Different systems of criteria and indicators have been empirically applied to assess the deliberative quality of the conversations (Borge & Santamarina, 2016; Friess & Eilders, 2014; Graham & Witschge, 2006; Kies, 2010; Klinger & Russmann, 2015; Steenbergen, Bächtiger, Spörndli, & Steiner, 2003; Stromer-Galley, 2007). Bächtiger et al. (2010) aimed to clarify that diversity, finding two main positions in the literature: Type I and Type II deliberation. The first one stems from Habermas's (1984) theory of communication and considers that a dialogue is deliberative if it is rational (based on logic, argumentation, and evidence), truthful, respectful, and oriented toward consensus or at least to mutual understanding. The second type is more empirically grounded and values alternative forms of communication (rhetoric and emotional discourse, storytelling, testimony) and includes other criteria such as plurality, diversity, and conflict (Bächtiger et al., 2010, pp. 43–46).

In the following paragraphs, we first explain briefly the institutional requirements and the desirable outcomes that should be considered to evaluate the deliberation in online platforms. Secondly, we will focus in more detail on the analysis of the communicative dimension.

Institutional or Structural Dimension and Outcomes of Deliberation

---

[3] The width of a conversation is the number of comments at any level of reply and the depth is the number of levels (Aragón et al., 2017, p. 5). Different combinations of these metrics are typically used to describe the structure of an online conversation.

The institutional or structural dimension refers to how online spaces should be structured to foster deliberation (Friess & Eilders, 2014, pp. 6, 15). Several institutional and technical characteristics are necessary for building a deliberative space, such as inclusion (Kies, 2010, pp. 42–44), asynchronous communication, content visibility, moderation, identity, perceived power of the communication spaces, relevant information, and horizontal interaction (Friess & Eilders, 2014, pp. 6–8). In the case of online discussions, several authors consider the tree-like arrangement of conversational threads formed by nested comments—similar to social websites like Reddit—to be the best interface for inducing deliberation (Aragón et al., 2017).

The criterion of inclusion means that all who are affected by and/or interested in the issues under discussion should be able to participate (Kies, 2010, p. 42). Thus, inclusion should be assessed by observing the technical characteristics of the online forum—that is, the ease of access in terms of connectivity and ICT skills—and discursive rules, such as moderation, registration, and identification, which should promote inclusive participation (Kies, 2010, p. 56).

In addition, following Friess and Eilders (2014, pp. 6–8), an asynchronous communication space is needed to allow participants to spend time reflecting and justifying their contributions. User content should appear immediately to motivate contributions and lower perceived entry barriers. Moderation is crucial to ensure deliberation in terms of civility and rationality and for promoting inclusive participation and good organization of the discussion (Steiner, 2012, pp. 256–258). Furthermore, empirical evidence shows that personal identification has positive effects on the deliberative quality of online debates (Fredheim, Moore, & Naughton, 2015), though other researchers also emphasize the importance of anonymity to push the number of participants (Friess & Eilders, 2014, p. 7). The perceived power of communication spaces refers to discussion spaces able to influence political outcomes. This condition encourages people to participate and be more deliberative (Friess & Eilders, 2014, p. 7), yet it may also lead to a competition-driven dynamic that negatively affects criteria such as reflexivity, sincerity, or empathy (Borge & Santamarina, 2016, p. 119).

The outcome dimension alludes to the results and impact of the deliberation, which can include improvements in civic attitudes and skills of the participants (Friess & Eilders, 2014, p. 10; Hendriks, Dryzek, & Hunold, 2007) or in the epistemic quality and legitimacy of decisions arisen from the conversations (Friess & Eilders, 2014, p. 10; Habermas, 1996).

In addition, Kies (2010, pp. 54–55) highlights the relevance of the external impact of the deliberative process outside the context of the debate. This means that decisions resulting from online forums should have an impact on public debates and political decisions and even shape binding norms to contribute to the participation of citizens and guide and scrutinize official decision-making processes (Dahlberg, 2007, p. 49; Hendriks et al., 2007).

Communicative Dimension

The communicative dimension refers to the deliberative attitude of participants (Kies, 2010, p. 42) and what the communication process should look like, mainly in relation to the reaction of participants to each other's ideas (Friess & Eilders, 2014, p. 8). As previously mentioned, there are two main strands of literature on how deliberative communication can be understood and assessed (Bächtiger et al., 2010). The first one follows the normative claims of deliberation theory, as stated in Habermas (1984), and regards deliberation as a systematic process of dialogue that is rational, inclusive, respectful, sincere, and where participants are willing to change their own opinions in light of better arguments, consensus being the final outcome (Friess & Eilders, 2014, p. 8). Here, one of the most crucial features is rationality in communication—that is, to state positions substantiated with logical arguments and empirical evidence. Within this strand, the most common deliberative criteria empirically tested are open and inclusive participation, justification (level and content), respect (toward others, groups, demands and arguments), and constructive politics or reflexivity (as a proxy to consensus) (Steenbergen et al., 2003; Steiner, 2012).

The second strand shifts away from the idea of purely rational dialogue toward alternative and more flexible forms of communication (Bächtiger et al., 2010, p. 33), acknowledges pluralism and difference, and relaxes the sincerity requirement (Bächtiger et al., 2010, pp. 43, 46). This second strand is more problem driven and empirically grounded, but still rooted in the deliberation's standards and procedures that should be achieved in the real world (Bächtiger et al., 2010, p. 45). As other studies focused on online deliberation (Friess & Eilders, 2014; Graham & Witschge, 2006; Kies, 2010; Klinger & Russmann, 2015), we consider that our study is closer to this second strand. We aim to measure the deliberativeness of spontaneous conversations on policy issues in a social online setting, adapting the common standards of justification, respect, inclusion,[4] and reflexivity.[5] We will include other criteria more clearly studied by this second strand of research such as reciprocity and plurality (Graham & Witschge, 2006; Gutmann & Thompson, 1996; Kies, 2010).

Especially in online settings, where interactions can be fast, short, and discontinuous, the assessment of reciprocity—that is, assessing whether participants are reacting and listening to previous comments—will show if there is a real discussion. For online media, a relevant indicator of reciprocity is the depth of the conversations (i.e., the existence of a chain of replies). Besides, as we aim to analyze a discussion on policy priorities, it is key to assess the level of plurality (i.e., the extent to which the debates host and confront all the relevant opinions on a specific topic; Kies, 2010, p. 53). Plurality can be assessed referring to disagreement and confrontation of different standpoints on policies (Kies, 2010, p.53) and should include a sufficient diversity such as different sociodemographic profiles and the presence of the actors with interests at stake (Klinger & Russmann, 2015; Young, 2002).

---

[4] The operationalization of the inclusion criteria in an online environment implies the institutional and technical requirements explained in the previous section, but also depends on the discourse equality of the communicative dimension—that is, participants should have equal opportunity to express any comment, and conversations should not be dominated by few participants (Friess & Eilders, 2014, p. 9; Kies, 2010, pp. 42–43).

[5] Participants should be able to change their initial opinion and preferences if they are persuaded by the force of the better argument (Bächtiger et al., 2010, p. 36; Kies, 2010, p. 48).

Leading theorists consider that conflict and disagreement act as a trigger for deliberation and help to sustain longer conversations (Fung, 2003; Gutmann & Thompson, 1996, pp. 73–80). Some empirical studies show that disagreement and different political positions induce more reciprocity and longer online conversations (Aragón et al., 2017; Balcells & Padró-Solanet, 2016).

As some authors are starting to acknowledge, the time dimension and the corresponding depth of the conversations have not been explicitly studied in relation to the unfolding of the deliberative criteria (Eilders & Esau, 2018). Yet the time dimension is implicit in the idea of deliberation itself, as any reason-giving exchange presupposes an ongoing process over time (e.g., Cohen, 1997, p. 72; Graham & Witschge, 2006, pp. 186, 195; Habermas, 1996, p. 306). Our approach is grounded on the assumption that time is required for the development of any genuine discussion. Any conversation is a process where interlocutors react to each other, and where any communicative action is a response to the previous action and affects the subsequent one. It seems reasonable, therefore, to expect variations of deliberative quality over time. We understand time in online conversations as the progressive development of a chain of interconnected messages, which is captured by the depth of a conversation thread. In this line, our contribution is not only to assess the deliberative quality of online conversations by a specific set of criteria and indicators but also to ascertain how the depth of online conversations affect the deliberative criteria.

Hypotheses

The operationalization of the deliberative criteria (discourse equality, reciprocity, justification, reflexivity, civility, plurality, and diversity) applied to the online conversation under study will be presented together with the data and methods (see Table 1). Overall, our research has two interrelated aims:

1. *To ascertain the degree of deliberation of this citizen proposal and the discussion it has generated, as well as to consider both the policy impact and the institutional design of the platform.*

2. *To analyze how the different criteria of deliberative quality evolve as participants interact in conversation.*

The first aim is more descriptive, whereas the second one is more exploratory and comprises different hypotheses.

Although some relevant studies have considered the different deliberative criteria as cohesively forming a common index (Klinger & Russmann, 2015; Steenbergen et al., 2003; Steiner, 2012), the attentive analysis of how conversations evolve over time reveals that deliberative traits may move toward different directions, and that it makes sense to examine them separately (Thompson, 2008, p. 512). In our study, we specifically understand time as the progressive development of a chain of interconnected messages, which is captured by the depth of a conversation thread. The dimension of depth, specifically in asynchronous communication, better measures the give and take of an evolving deliberation than does simply considering the moment in time in which a given comment is posted or how long the conversation last.

Considering the dimension of time allows one to assess different degrees of deliberative quality. Actually, differences in quality can be observed not only by comparing deliberative criteria between them but also their ongoing evolution through interaction between participants. As widely discussed in the literature, multiple factors can affect the development of deliberative quality, ranging from technical design to the characteristics of the participants or the topic of discussion, inter alia. Our purpose here is not to discern the specific causes that explain the differences in deliberative quality—which would obviously require a research design oriented toward causation—but rather to describe different patterns of evolution of deliberative criteria, given the particularities of the case under analysis. That can contribute to a better understating of deliberation dynamics in online digital platforms such as Decidim.

We consider several possible hypotheses that relate the different indicators of deliberative quality with the development of conversations over time (see Figure 1).

H0:   *Assumes no relation between the development of the conversation and the criteria of deliberative quality. This null hypothesis would consider deliberative criteria to be a stable phenomenon with no variation over time.*

H1:   *Assumes indicators of deliberative quality to increase as conversations develop. This hypothesis would imply that certain quality elements are especially activated in the heat of debate. It could also indicate a kind of virtuous circle; the existence of deliberative quality would progressively lead to better quality. This would be the expected relation of criteria such as reciprocity, as continuous interaction in public is likely to encourage a more personal and direct way of communication.*

H2:   *On the contrary, H2 assumes deliberative quality indicators to decrease as conversation further develops. This hypothesis would show instead a reduction of quality that could be explained by a plurality of different factors, such as a certain fatigue due to persistent conflict and strong disagreement. For example, this could be the case of criteria such as justification if arguments are presented in a more complete form at the beginning of the conversation and then progressively fade away.*

H3:   *H3 acknowledges that H1 and H2 are an oversimplification of the complex dynamics we can expect within an actual conversation, as they assume a monotonous linear convergence or divergence among interlocutors. However, H3 improves the complexity and realism of previous hypothesis by supposing a curvilinear path in the development of the conversation. Either the conversation process starts with a path toward mutual understanding and respect that is broken at some moment (H3.1), or, after a starting point of conflict and divergence, the conversation ends up with better mutual understanding (H3.2). Therefore, H3 is a combination of the two previous hypotheses that could fit better with the complexity of conversations as it assumes no linearity. H3.1 shows that quality grows until a certain point of fatigue or exhaustion, from which it starts to decrease. H3.2 instead shows the reverse pattern, that quality decreases over time, but is reactivated at the end of the conversation.*
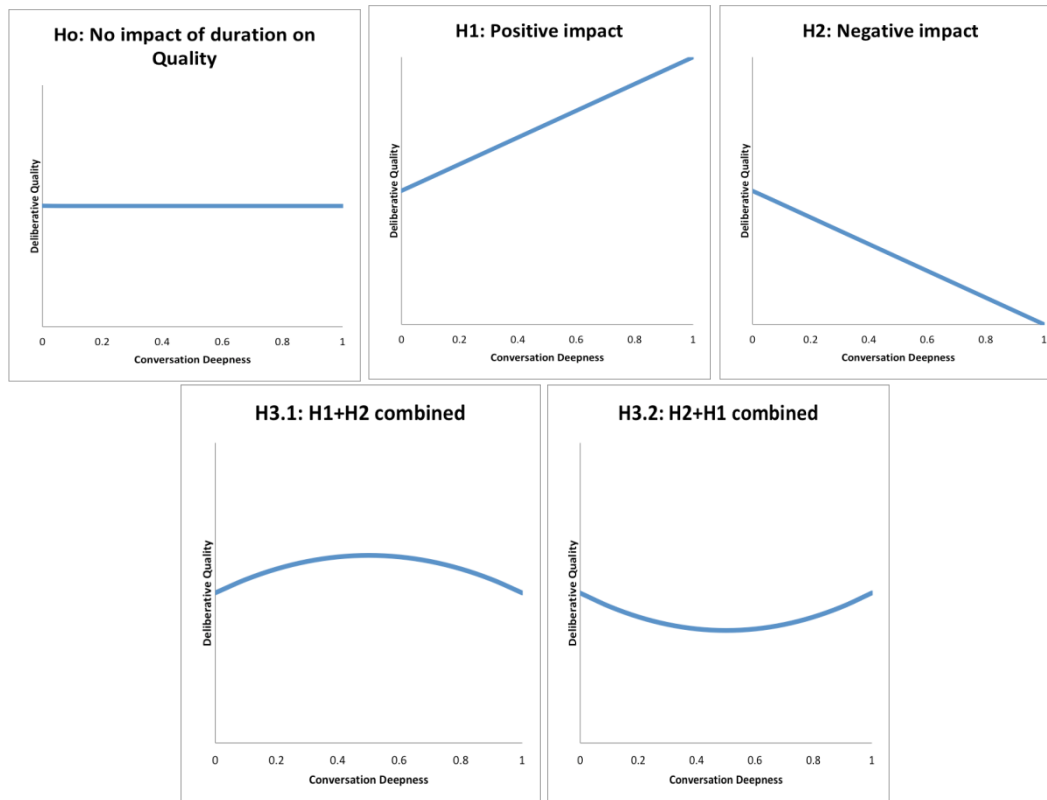
Figure 1. Hypotheses modeling the relation between deliberative quality
(vertical axis) and development of conversation, in terms of depth (horizontal axis).

The New Participatory Platform Decidim and the Barcelona Case Study

Barcelona constitutes a good laboratory for citizen participation for several reasons. First, the city and its surroundings encompass a dynamic urban area with strong civic networks, a rich associative life, and a widespread use of digital technologies.[6] Second, it has a long record of participatory experiences and citizen mobilizations, most recently connected with the cycle of protest triggered by the economic crisis and the conflict over Catalonia's independence. Third, since the recovery of democracy in the last quarter of the 20th century, participation has been a paramount concern for the successive local governments in Barcelona, which have invested many resources in participatory policy making. This has been intensified after the 2015 local elections, with the arrival to power of a new political party (Barcelona en Comú), mostly composed of civil society activists with a strong commitment toward citizen participation. The mayor herself, Ada Colau, was a leader of the anti-eviction movement.

---

[6] Barcelona usually appears in advanced positions in the international rankings on networked and innovative cities (European Capital of Innovation in 2014 or number 13 in the Networked Society Index in 2016).

Barcelona's Strategic City Planning was the first local government's action to give citizens the possibility of defining from below policy priorities. This first participatory process was hosted in the newly developed Decidim platform, combining both digital and face-to-face participation. In the Strategic City Planning, citizens were able to make policy proposals through the digital platform, and each proposal could be openly discussed in a forum and voted on by other citizens who had previously registered. The Planning had in total 10,860 proposals (9,560 initiated by citizens), 18,192 comments, and 25,435 online participants. The presentation of the proposals and the corresponding debates took place between February 1 and mid-April 2016. They were not conducted by moderators, but relied on spontaneous participation. Even though registration was necessary, citizens could use nicknames to make communications anonymous.

The most commented policy proposal was granting new licenses for tourist apartments. Despite the number of comments and supports, the proposal was finally rejected because it was considered to go against the political line of the local government, one of whose main priorities has been to regulate the tourist boom. The debate on tourism is at this moment highly controversial in Barcelona. Though an important source of income, the negative externalities of mass tourism—such as gentrification—have become a troublesome problem for locals. The discussion on the model of tourism has become a matter of intense political debate. The opening of the platform was probably seen as an opportunity to bypass or influence decision makers before a public audience and put the issue of new licenses into the political agenda. The online debate reflects the different interests at stake—namely, the confrontation between property owners and local neighbors, and their reasons for favoring or blocking the proposal.

Operationalization, Data, and Methods

First, we ascertained how the online platform was structured and technically organized to foster deliberation, and checked whether the results of the debate were accepted by the municipal government. Second, regarding the communicative dimension, we carried out content and network analyses of the conversation raised around the proposal to grant new licenses for tourist apartments. This proposal was registered on February, 28, 2016, and the conversation ended up on April, 4, 2016. It received 335 comments, 198 votes, and 72 citizens took part in the conversation. Among the 20 most commented proposals in the Strategic City Planning, three focused on tourist lodging and one on Strategic Tourism Planning. Figures 2a and 2b show an extract of the conversation.

Figure 2a. Citizen proposal on granting new licenses for tourist apartments.

Figure 2b. Part of the conversation raised about the proposal on granting new licenses for tourist apartments.

Following the deliberative criteria compiled in the theoretical framework, in Table 1 we present the operationalization of the indicators for each criterion. As recommended for any empirical analysis of deliberation (Kies, 2010, p. 55; Steiner, 2012, p. 13), the operationalization is adapted to the context and type of conversation and space.

Table 1. Meaning and Operationalization of Deliberative Criteria.[7]

Delib. criteria	Operationalization	Indicators
Discourse equality	Analysis of discursive concentration by few participants and whether this concentration leads to control of the debate.	▪ Number of comments per participant, Mean and *SD* ▪ Number of participants that concentrate more comments ▪ Gini Index[8]
Reciprocity	Proportion of comments that are part of a thread versus those that initiate a thread but without followers, and by measuring the extent to which comments take into consideration opinions of a preceding posting.	*REC Index*: 1. Literal mentioning of the name of the participant to whom the comment is referring. 2. Literal citation in quotes of previous comments or parts of the comments 3. Questions that refer to previous comments or are rhetorical or an interpellation *Other*: ▪ Number of comments by level of depth.
Justification	Observing whether opinions and proposals are justified and how complex the justifications are. Analysis of whether the justification's arguments are internal (based on personal viewpoints, values, of situation) or external (based on facts, figures, comparisons, proposals, or links to other information).	*JUST Index*: 1. External justification based on data or on the content of laws and regulations. 2. External justification based on facts or statements as if they were facts. 3. External justification based on comparisons or similes. 4. External justification referring to the common good or interest, justice, or rights or to a collective or social class. 5. External justification based on links to websites, videos, or different information sources. 6. External justification based on solutions or proposals. 7. Internal justification referring to the personal experience or situation. 8. Internal justification referring to own values, rights, ideologies, or viewpoints. *Other*: ▪ Very short comments without justification. ▪ Number of characters of the comments.

[7] Based on Kies (2010, pp. 42, 52–57).

[8] The Gini index is used here as a measurement of inequality in the production of comments. Zero means maximum equality, where every participant contributes with the same number of comments. The closer the index gets to one, the higher the concentration of comments among a few participants.

Reflexivity	Notifying visible instances of opinion changes, conflict resolutions, moderation of the conversational tone, or solutions and conclusions based on comments from other participants.	*REF Index*: 1. Opinion changes. 2. Conflicts resolution. 3. Moderation of the conversational tone, conciliation. 4. Conclusions based on previous comments. 5. Solutions based on proposals from other participants.
Civility	Counting the cases of disrespect, accusations, insults, and negative ironies or jokes. Also collecting acknowledgements and positive remarks.	*Positive remarks*: Thanks, acknowledgement, admiration, enthusiastic agreement. *Incivility or disrespect index*: 1. Accusations. 2. Ironies. 3. Jokes. 4. Insults. 5. Ridicule. *Other*: ▪ Repeated comments.
Plurality and diversity	Analyzing the degree to which the debates refer to different political ideologies and whether there is disagreement and conflicting standpoints. The registration system, if it is not anonymous, could give information about gender. It is also possible to identify the language of the comments.	*Plurality of viewpoints*: ▪ Approval/disapproval of the initiating proposal. ▪ Conflictual comment that is an opposing comment to the previous one. ▪ In favor or against local government position or regulations on the issue at stake *Diversity*: ▪ Proportion of women and men. ▪ Proportion of comments on different languages (Catalan, Spanish).

The coding process developed in the following phases. First, four coders carried out a previous codification of 50 comments to test and agree on the coding dictionary. Second, the total coding of the 335 comments was carried out by three coders. After the coding was completed, intercoder reliability through Krippendorff's alpha was calculated. The results were not satisfactory for 20 indicators (of the total 28). So, third, we decided to clarify the description of the indicators, improve the instructions for codification, add more information to the codebook and finally recodify the values for the 20 indicators. Eventually, eight of them were dropped because of lack of variability. That was the case of the majority of the indicators of reflexivity and two indicators of incivility. In controversial political discussions, it is difficult to achieve a high level of reflexivity, which implies changes of the initial positions and efforts to arrive to a common agreement (Hendriks et al., 2007, pp. 369–370; Wojcieszak, 2011, pp. 328–330). In addition, the indicator of external justification regarding the common or collective good, interests, or rights was removed because of the difficulty for coders to agree. Finally, after the recodification, we have in total 20 indicators or variables (see Table 2 in 'Analysis and Discussion' section), most of them aggregated into different indexes, with reasonably good levels of intercoder reliability (with Krippendorff's alphas ranging between 0.713 and 0.922).

To test the hypotheses, we ran several regression analyses between the indicators of deliberative quality and the evolution of conversations over time. To capture the temporal evolution of conversations, we first identified the different conversation threads that composed the whole discussion. By conversation thread we mean any initial comment that directly responds to the proposal and starts a line of discussion, together with all the subsequent comments that are derived from it. Second, we have identified the position of each comment in the thread, by specifying its level of depth (that is to say, the layer it belongs to). Depth can be regarded as a good indicator of development because each level of depth represents a step further in the chain of replies from previous comments; thereby, it captures the natural give and take of conversations. To make it comparable across different conversation threads (because some can be very short, but others very long), the variable "depth" has been rescaled to range between 0 and 1; where 0 means that the comment is the beginning of a conversation thread, 1 means that it is the ending, and all the other values in between show intermediate developments of the conversation thread.

Analysis and Discussion

Institutional Design and Impact of the Deliberation

The institutional design refers to the main structural and technological conditions that foster deliberation. The examination of the functionalities of Decidim confirms that most of the structural requirements for deliberation are satisfied. The platform allows asynchronous participation (i.e., users can contribute to their post at any moment), which provides more time for reflection and justification. User content appears immediately, allowing horizontal interaction by either commenting on other user's threads and comments, or voting on contributions. The platform administrator acts as a nonintrusive moderator by posting a link to similar proposals or announcing the merge of the proposal (but not of the conversations) with others in a single action plan. Participants can denounce a comment in cases of disrespect, attacks, or insults by using a flag tag, which can be considered as a form of self-moderation. User registration requires only a username (or nickname), a password, and an e-mail address. With registration it is possible to post a comment, but voting requires the postal address and ID card.

We have detected some problems affecting the deliberative quality of the conversation that might have been solved with better moderation and a more organized structure around citizens' proposals. Dozens of proposals were related or belonged to the same issue, but were not properly merged into a single conversation or a general proposal. In addition, the debate was not structured around relevant information on the issue supplied by official administrators, nor were there facilitators aimed at helping participants build an inclusive discussion. Deliberation was simply expected to emerge in a decentralized and spontaneous way. In fact, the debate entirely depended on the information provided by the participants themselves, who obviously had personal—and often narrow—interests on the matter. This can be problematic when, as it happened, there is a lack of representation of all the interests at stake in the discussed issue. As could be deduced by the arguments and personal experiences, the debate mostly involved the confrontation between small owners of flats and local residents, which resulted in discourse concentration by some few and sometimes intransigent participants.

The local government committed itself publicly to accept proposals that received the most votes, but the most commented ones were also considered. In total, 75% of the proposals presented for the Strategic City Planning were accepted (Ajuntament de Barcelona, 2017). However, the local government reserved the right to filter and reject proposals that were not in line with their political priorities. That is what happened with the proposal under study, which was rejected with very short feedback: "The local government is not working on this line" (see Figure 2). The lack of feedback can cause frustration among participants and defendants of the proposal., The local government was not able to send a proper response to all 9,560 citizen proposals. Indeed, online debates about the Strategic City Planning took place from February 1 to April, 15, 2016, only six months after the electoral victory. Currently, the feedback functionality at the platform has improved, and more civil servants work in this important task (Barandiaran & Romero, 2017).

The Deliberative Quality

The success of online debate on new licenses for tourist apartments is due to its controversial nature and the involvement of a few users—who have personal interests on the issue—that have triggered a noteworthy cascade of comments. Comparatively, this is an exceptional case, because the majority of proposals (51.7%) on the platform have not generated any debate at all. If stakeholders have decided to take part in this particular debate, it is because they have considered the issue to be relevant and have perceived that their participation could make the difference between success and failure of the proposal.

The debate has mostly been dominated by few users who have posted the majority of comments (see Figure 3). The level of discourse equality is very low, as can be seen in all indicators that measure discourse equality. For example, only four individuals, who are slightly above 5% of the total number of users, concentrate 65% of the posts, with a Gini index of .71; the average number of posts written by users is five, with a standard deviation of 14 and a median of one (which clearly shows the distribution asymmetry).

Despite the uneven participation, as shown in Table 2, the indicators of plurality reveal that the debate has captured the attention of stakeholders with antagonistic interests (66% of the comments are in favor of the proposal, and 31% are against it). The confrontation between defenders and opponents of the proposal—though asymmetrical—has been balanced (69% of comments imply the interaction between individuals who hold opposite opinions). The structure of the conversation tree shows that most threads are actually a chain of replies that successively alternate for and against positions (see Figure 4). However, there is a participation bias in favor of those who support the proposal. That might be because participation in the platform could be seen as an opportunity for small local property owners who had no other means to protest against the local government's policy of limiting tourist licenses. Thereby, participation in the debate was a way to make visible their discontent by publicly exposing their views. That has triggered the opposition of several local neighbors who identified themselves as living in the areas now more affected by the negative externalities of mass tourism.
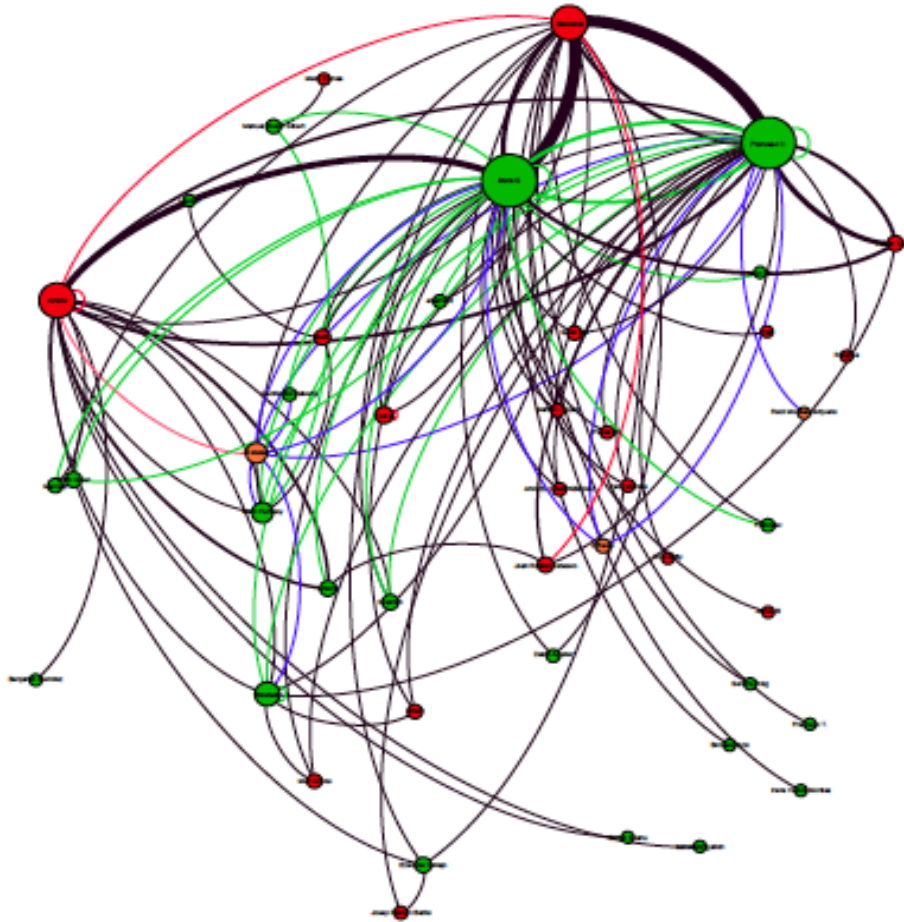
Figure 3. Network of participants (only comments above level 1 of depth). The nodes' size varies according to the level of degree. Green nodes are participants who favor the proposal; red nodes are against it; orange means undefined positions. Edges are scaled by intensity of exchanges; black lines mean interaction between opponents; red/green lines mean interaction between like-minded users; blue lines imply interactions with undefined users.
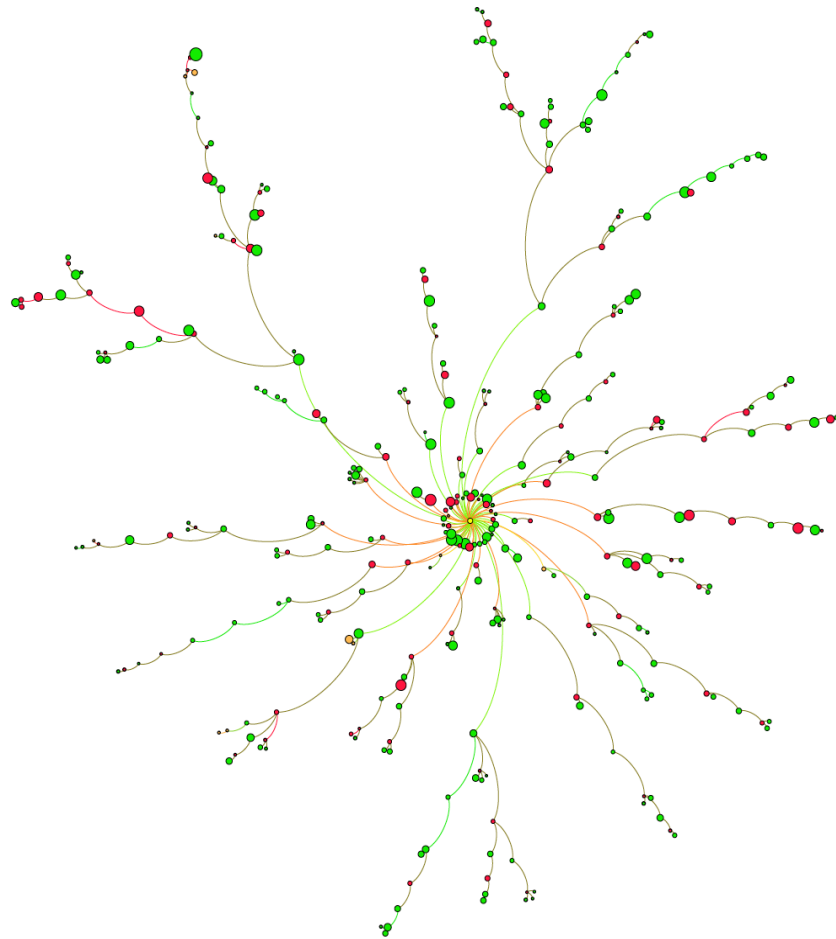
Figure 4. Radial tree of the debate. Green nodes are comments in favor of the proposal; red nodes are comments against the proposal (orange ones are undefined cases). The central node is the proposal. Nodes (N = 335) are scaled according to the number of characters of each post.

The proportion of women is higher (41.5% of women) than in other online political fora where men usually represent two-thirds of participants (Borge & Santamarina, 2016, p. 119). As for the language used by participants, the initial proposal was in Catalan, and 70% of the comments were in Catalan and 30% in Spanish. Frequently, the use of one or the other language depends on the language of the previous comment, as usually happens in face-to-face conversations.

**Table 2. Summary of Variables by Dimension of Deliberative Quality
(Percentage of Appearances).**

Dimension	Variables	Percentage	Krippendorff's alpha	Index*
Reciprocity (0–3)	REC1 mentions	32.8	0.73	0.68
	REC2 quotes	7.8	0.916	(0.81)
	REC3 questions	27.8	0.791	
Justification (0–7)	JUS1 EXT data	13.4	0.922	1.53
	JUS2 EXT facts	62.4	0.715	(1.06)
	JUS3 EXT comparisons	23.3	0.713	
	JUS4 EXT links	4.2	0.776	
	JUS5 EXT proposals	22.4	0.803	
	JUS6 INT personal experience	19.1	0.732	
	JUS7 INT values	8.7	0.861	
Reflexivity (0–1)	REF1 moderation	3.3	0.963	0.03
				(0.18)
Positive remarks (0–1)	Thanks, recognition	9.0	0.708	0.09
				(0.28)
Incivility (0–3)	INCIV1 accusation	34.3	0.765	0.49
	INCIV2 irony	10.8	0.858	(0.61)
	INCIV3 insults	3.9	0.906	
	Repeated comments**	13.7	0.822	
Plurality	PLU1 approval/disapproval	66 (approval) 31 (disapproval) 3 (missing)	0.848	-
	PLU2 conflict***	69.2	Based on PLU1	
Diversity	DIV1 gender	41.5 (women) 53.4 (men) 5.1 (n/a)	0.917	-
	DIV2 language	70 (Catalan) 30 (Spanish)	0.82	

Note. The last column includes a summary index of each dimension, excluding repeated comments, plurality, and diversity.
*Mean value and standard deviation (between parentheses).
** Not included in the index of incivility.
*** Conflict means an opposite reply to a previous comment, but based on their opinion on the proposal (PLU1).

The interaction and attention between users is captured by the reciprocity dimension, with direct mentions between users, the use of direct questions, and literal quotes of other users' expressions or

statements. As shown in the distribution of comments by level of interaction (see Figure 5), only 67 comments (20%) belong to the first level of interaction (i.e., they are an immediate reply to the initial proposal). Remarkably, 203 comments (60.6%) belong to Levels 3–9 in the chain of replies (with the maximum depth in the conversation at Level 9), which indicates that reciprocity between participants is relatively frequent.
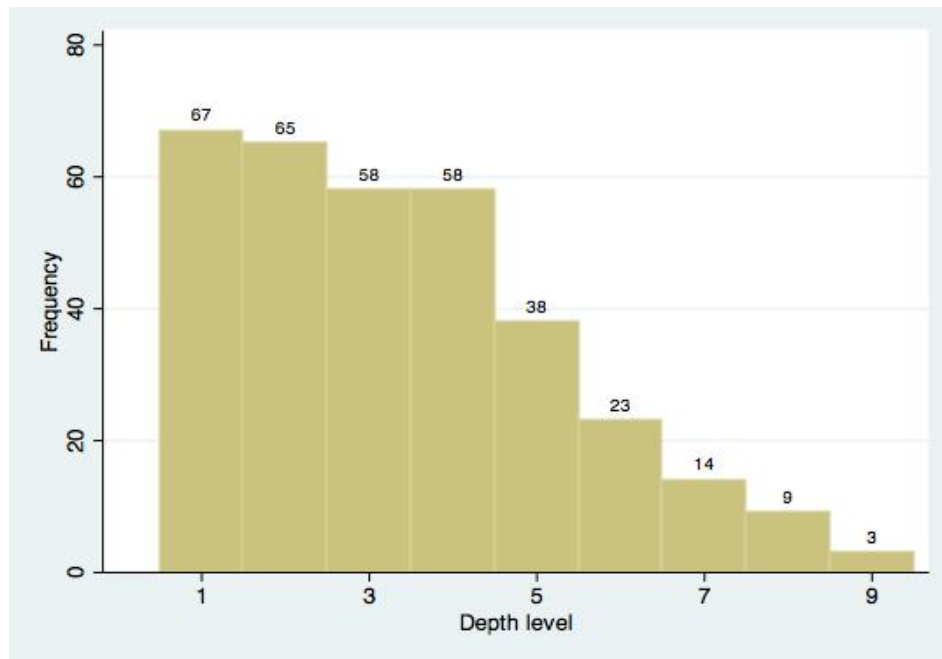
Figure 5. Frequency of comments by level of interaction or depth.

As shown in Table 2, the debate is characterized by a real exchange of arguments, which combines a repertoire of different forms of justification, such as the description of facts (62.4%), comparisons (23.3%), solutions or proposals (22.4%), references to personal experience (19.1%), and to a lesser extent, own ideological values (8.7%) and links (4.2%). These relatively high percentages show that there has been a serious attempt to provide evidence for supporting one's point of view or attacking the weaknesses of opposing arguments. It is also remarkable that 1 of 5 comments suggest solutions or proposals about the issue under discussion, which shows also the propositive nature of the debate with the discussion of different alternatives. In addition, the average number of characters of the comments is fairly high (307), representing approximately four sentences within the space provided by the interface (see Figure 2), the median being 253 characters (around three sentences), and the standard deviation 238.

The levels of reflexivity (in terms of moderation of the tone of the debate) and positive remarks are comparatively more modest, with percentages below 10%. The high levels of interaction and argumentation are not incompatible with elements of incivility or disrespect, such as accusations (34.3%), irony (10.8%), or even insults (3.9%). Even though these elements cannot be seen as a sign of quality, it

is not obvious that they necessarily play a negative role in deliberation (Young, 2002, p. 49). For instance, when combined with the presentation and development of arguments, they can increase the degree of engagement in conversation, by inviting users to reciprocate and respond (Balcells & Padró-Solanet, 2016). However, they may become problematic if they come at the price of replacing the exchange of arguments tout court. Indeed, the number of repeated comments are more frequent when part of the dialogue has already taken place (72% of the repeated comments appear in the second half of the conversation). Most of them belong to the three most active participants in the conversation (see Figure 3), who also have more extended comments.

The Evolution of Deliberative Quality

We have carried out several regression analyses (ordinary least squares; OLS) to test the relation between the development of conversations over time, in terms of depth, and five indicators of deliberative quality (reciprocity, justification, reflexivity, positive remarks, and incivility; see Table 3). The model includes the quadratic transformation of the variable depth given that, theoretically, the relation could be nonlinear, as expressed in H3. Two indicators of plurality are included in the model as influencing the other deliberative criteria because, as explained in the theoretical framework, disagreement (PLU1) and conflict (PLU2) may help to trigger longer conversations. The length of a comment—measured in number of characters—is also included because it could be logically associated with most deliberative criteria such as justification, reciprocity, and reflexivity (Kies, 2010, p. 47).

As stated in the hypotheses, the quality of deliberation varies as conversation threads unfold, but, remarkably, indicators evolve differently. The regression analyses (see Table 3 and Figure 6) show that the level of justification decreases as conversations go deeper, whereas the levels of reciprocity and incivility both become more important over time. In both cases, the pattern is more similar to what was stated in H3, because the relation is not completely linear (as the coefficient of the quadratic transformation of the variable "depth" shows). The depth of conversations has no statistically significant coefficients in the case of reflexivity and positive remarks, showing that there is no relation between them.

As explained in the hypotheses section, the five deliberative dimensions may theoretically have an independent evolution over time, so we have tested their reciprocal independence with our data. The empirical tests of independence show no significant linear relations among them. Two of the bivariate Pearson correlations between the dimensions are just below 0.2, and the others are even lower. A principal components analysis shows that justification strongly dominates the first component (24% of the variance), positive remarks and reflexivity load positively on the second component (24% of the variance), whereas incivility and reciprocity load negatively on this second component. Altogether, these tests confirm that the five deliberative dimensions capture different traits of the conversation.

Table 3. Regression Models (OLS) between the main Indicators of Deliberative Quality and the Development of Conversations (in terms of depth).

	Reciprocity		Justification		Reflexivity*		Positive remarks*		Incivility	
	Coef.	$p > \|t\|$	Coef.	$p > \|t\|$	Coef.	$p > \|t\|$	Coef.	$p > \|t\|$	Coef.	$p > \|t\|$
Depth	1.486	.005	−1.250	.027	0.125	.325	0.045	.791	1.380	.001
Depth^2	−1.025	.032	0.936	.065	−0.093	.418	−0.086	.574	−1.041	.005
# characters	0.001	.000	0.003	.000	0.000	.367	0.000	.406	0.000	.216
PLU1 approv.	0.014	.893	0.201	.079	−0.044	.088	0.006	.853	0.020	.810
PLU2 conflict	0.182	.094	0.315	.006	0.006	.816	−0.156	.000	0.107	.202
Constant	−0.071	.693	0.701	.000	0.018	.671	0.212	.000	0.035	.800
R^2	.089		.364		.016		.073		.054	
$N = 291$**										

*Although positive remarks and reflexivity are dichotomous, the estimation of a probability regression model is unproblematic given that, because of the inclusion of the nonlinear quadratic term, the values do not exceed the theoretical range (0, 1), as shown in Figure 6 (Harrell, 2015).

**The final N is slightly below the total number of comments (335) because comments on the initial proposal that have not started a line of discussion have been removed from the models

In Figure 6, we show the predictive values of the deliberative quality indicators at different depth levels, holding the rest of the independent variables at their average values.

These findings indicate that conversation threads in this debate are developed through different phases, which could be summarized as follows. Initially, users try to present convincing arguments to justify their positions. At this initial moment, most of the efforts are focused on developing persuasive arguments and providing enough evidence to support them. However, afterward there is a certain general effect of exhaustion through interaction. This can be partly explained because conversations commonly start as a statement with a more or less complete justification of the argument, and as the discussion develops, they then become more focused on specific details. In addition, the incompatibility of viewpoints and the lack of a common ground can also diminish the emphasis on justification, by transforming the conversation into a more personal and less rational exchange of views in the form of irony, personal attacks, and accusations reciprocally responded to by each side. This phase of degradation can lead to an isolation of the conversation, making it less appealing for other users to take part in the debate and, as a consequence, reducing the visibility and the necessity to present widely acceptable arguments. Overparticipation by a few highly active users does not help to keep personal confrontations out of the debate. Reiterative interactions between the same users sometimes lead to a repetition of arguments, even a literal copy and paste of previous posts, transforming the debate into a sort of dialogue of the deaf.
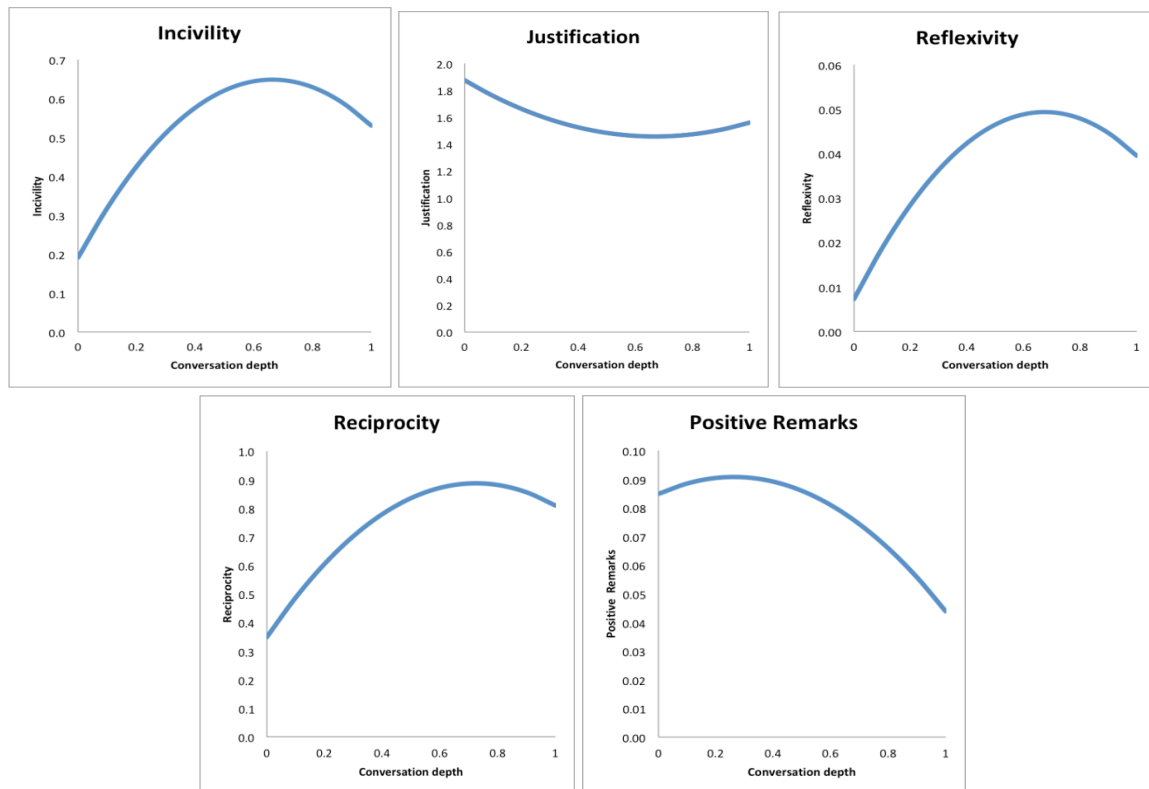
*Figure 6. Evolution of deliberative quality indicators through
depth according to regression models.*

Overall, these findings show that spontaneous deliberation is possible, but they also reveal the difficulties of ensuring and preserving the quality of deliberation over time. The discussion mostly tended toward exposing the reasons for or against the proposal, and, despite real attempts to exchange arguments, there were at the end few concessions to either side. Two elements have probably helped to produce this result. On the one hand, the possibility of voting on the proposal while discussing it generates incentives for campaigning to add or subtract supports, but not for abandoning maximalist positions and compromising. On the other hand, stakeholders have probably perceived through successive interactions that interests were so antagonistically opposed that no possible common ground was achievable, and that it was not worth the struggle for a mutually agreed on solution. Thus, continuous interaction has probably served to reinforce initial positions instead of moderating conflict and reconciling the different viewpoints. At the very least, conversations may have helped to voice the different interests at stake and make visible the diversity of stances and claims.

Conclusions

This article shows the potential of participatory platforms like Decidim to become a space for public deliberation. We find that online conversations can achieve minimal criteria of deliberative quality, though we also identify some problems and pitfalls that could undermine them. Two main contributions are highlighted in our research. First, we find that indicators of deliberative quality are not necessarily stable, but may evolve differently over time. Bridging the gap between the literature on deliberative quality and the analysis of conversation structures, we show how deliberative quality indicators vary as conversation threads move forward. Second, we find that deliberative criteria are not a monolithic index, but they can evolve differently and even in opposite directions. Overall, these two findings help to complete the empirical literature on deliberative quality, by stressing the time-dimension perspective offered by the idea of depth. Although a matter of interest by the computing-based literature on conversation structures due to the characteristics of online systems—where all data are registered and visible—that point has been insufficiently explored in empirical analysis on deliberative criteria.

Genuine citizen deliberation can emerge spontaneously in online spaces habilitated for debate—as happened in the case under analysis—but it can be easily eroded, too, if there are no favorable conditions. For instance, we find that the justification of arguments decreases as conversation threads develop, whereas certain elements such as incivility gain further relevance. We find high levels of confrontation between opposing views, which temporarily increase some features of the deliberative quality, such as reciprocity. But in the long run, as conversations go deeper, the lack of common ground and moderation, and the discourse concentration by few participants, has negatively affected the quality of deliberation (lack of civility, repetition of comments, and no reflexivity). The identification of these trends may offer useful information for institutions to understand the dynamics of conversations and provide the right conditions for keeping online debates within good standards of deliberation. The deliberative erosion found in this conversation can be partly explained because of some problems identified in the institutional design. Even though the platform under analysis seems to work well for enabling citizens to voice their claims, certain additional features are needed if conversations are expected to more specifically aim toward higher standards of deliberation. Social-media-like structure, as the one used in the Decidim platform, can facilitate interactions among users in a very intuitive way, and thus might provide a good basis for communication. However, certain requirements—such as active moderation, previous information, some basic regulatory norms, representation of all stakeholders, and better voting and aggregation procedures—should be seriously taken into account if deliberation is to be less vulnerable to the randomness of spontaneity or to the appropriation of few users and advocate groups.

References

Ajuntament de Barcelona. (2017). *Programa d'Actuació Municipal 2016–2019* [Strategic City Planning 2016–2019]. Retrieved from https://decidim-barcelona-new.s3.amazonaws.com/decidim-barcelona/uploads/decidim/attachment/file/10/00.pdf

Aragón, P., Kaltenbrunner, A., Calleja-López, A., Pereira, A., Monterde, A., Barandiaran, X. E., & Gómez, V. (2017). Deliberative platform design: The case study of the online discussions in Decidim Barcelona. In G. L. Ciampaglia, A. Mashhadi, & T. Yasseri (Eds.), *Social informatics* (pp. 277–287). Cham, Switzerland: Springer International.

Bächtiger, A., Niemeyer, S., Neblo, M., Steenbergen, M. R., & Steiner, J. (2010). Disentangling diversity in deliberative democracy: Competing theories, their blind spots and complementarities. *Journal of Political Philosophy*, *18*(1), 32–63. doi:10.1111/j.1467-9760.2009.00342.x

Balcells, J., & Padró-Solanet, A. (2016). Tweeting on Catalonia's independence: The dynamics of political discussion and group polarisation. *Medijske Studije*, *7*(14), 124–141. doi:10.20901/ms.7.14.9

Barandiaran, X. E., & Romero, C. (2017). *Decidim's functions and features.* Retrieved from https://docs.decidim.org/features/en/general-description/

Borge, R., & Santamarina, E. (2016). From protest to political parties: Online deliberation in new parties in Spain. *Medijske Studije*, *7*(14), 104–122. doi:10.20901/ms.7.14.8

Cohen, J. (1997). Deliberation and democratic legitimacy. In J. Bohman & W. Rehg (Eds.), *Deliberative democracy: Essays on reason and politics* (pp. 67–91). Cambridge, MA: MIT Press.

Dahlberg, L. (2007). The Internet, deliberative democracy, and power: Radicalizing the public sphere. *International Journal of Media & Cultural Politics*, *3*(1), 47–64. doi:10.1386/macp.3.1.47_1

Dahlgren, P. (2005). The Internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication*, *22*, 147–162.

Eilders, C., & Esau, K. (2018, November 8–9). *It takes more than good arguments! Integrating approaches on deliberation, social influence and social choice.* Paper presented at the Deliberative Quality of Communication Conference, Mannheim, Germany.

Fredheim, R., Moore, A., & Naughton, J. (2015). Anonymity and online commenting: The broken windows effect and the end of drive-by commenting. *Proceedings of the ACM Web Science Conference*, *11*, 1–8. doi:10.1145/2786451.2786459

Friess, D., & Eilders, C. (2014). *Analyzing crowd discussion towards a more complete model to measure and explain online deliberation*. Paper presented at the Internet, Politics and Policy, Oxford

Internet Institute, Oxford, UK. Retrieved from
http://ipp.oii.ox.ac.uk/sites/ipp/files/documents/IPP2014_Friess.pdf

Fung, A. (2003). Recipes for public spheres: Eight institutional design choices and their consequences. *Journal of Political Philosophy*, *11*(3), 338–367. doi:10.1111/1467-9760.00181

Gómez, V., Kaltenbrunner, A., & López, V. (2008). Statistical analysis of the social network and discussion threads in Slashdot. *Proceedings of the 17th international conference on World Wide Web, ACM* (pp. 645–654). Retrieved from http://ra.ethz.ch/CDstore/www2008/www2008.org/papers/pdf/p645-gomezA.pdf

Gonzalez-Bailon, S., Kaltenbrunner, A., & Blanchs, R. E. (2010). The structure of political discussion networks: A model for the analysis of online deliberation. *Journal of Information Technology*, *25*, 1–14. doi:10.1057/jit.2010.2

Graham, T., & Witschge, T. (2006). In search of online deliberation: Towards a new method for examining the quality of online discussions. *Communications*, *28*(2), 173–204. doi:10.1515/comm.2003.012

Gutmann, A., & Thompson, D. F. (1996). *Democracy and disagreement*. Cambridge, MA: Belknap Press.

Habermas, J. (1984). *The theory of communicative action: Reason and the rationalization of society* (T. McCarthy, Trans.). Boston, MA: Beacon Press.

Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy* (W. Rehg, Trans.). Cambridge, MA: MIT Press.

Harrell, F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). New York, NY: Springer.

Hendriks, C. M., Dryzek, J. S., & Hunold, C. (2007). Turning up the heat: Partisanship in deliberative innovation. *Political Studies*, *55*(2), 362–383. doi:10.1111/j.1467-9248.2007.00667.x

Kies, R. (2010). *Promises and limits of web-deliberation*. New York, NY: Springer.

Klinger, U., & Russmann, U. (2015). The sociodemographics of political public deliberation: Measuring deliberative quality in different user groups. *Communications*, *40*(4), 471–484. doi:10.1515/commun-2015-0017

Steenbergen, M. R., Bächtiger, A., Spörndli, M., & Steiner, J. (2003). Measuring political deliberation: A discourse quality index. *Comparative European Politics*, *1*(1), 21–48. doi:10.1057/palgrave.cep.6110002

Steiner, J. (2012). *The foundations of deliberative democracy: Empirical research and normative implications*. Cambridge, UK: Cambridge University Press.

Stromer-Galley, J. (2007). Measuring deliberation's content: A coding scheme. *Journal of Public Deliberation*, *3*(1), 1–37.

Thompson, D. F. (2008). Deliberative democratic theory and empirical political science. *Annual Review of Political Science*, *11*(1), 497–520. doi:10.1146/annurev.polisci.11.081306.070555

Wojcieszak, M. (2011). When deliberation divides: Processes underlying mobilization to collective action. *Communication Monographs*, *78*(3), 324–346. doi:10.1080/03637751.2011.589459

Young, I. M. (2002). *Inclusion and democracy*. Oxford, UK: Oxford University Press.