Examining the Quality of Crowdsourced Deliberation: Respect, Reciprocity and Lack of Common-Good Orientation

Tanja Aitamurto
Stanford University
Stanford, CA 94305, USA
tanjaa@stanford.edu

Jorge Saldivar
Catholic University "Nuestra Señora de la Asunción"
Asunción, Paraguay
jorge.saldivar@uc.edu.py

CHI'17 Extended Abstracts, May 06-11, 2017, Denver, CO, USA.
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3053248

Abstract
We examine deliberative quality of crowdsourced deliberation in this paper. Analyzing data from two crowdsourced policy-making processes, we found a good quality deliberation with respect, reciprocity, and storytelling according to the standards in the theory of deliberative democracy. We identified a group of super-deliberators, whose deliberation was above the average, and low-quality deliberators, whose deliberation was below the average. The findings show that even when crowdsourced policymaking was not designed for deliberation, it can facilitate a fairly high-quality democratic deliberation.

Author Keywords
Crowdsourcing; crowdlaw; crowdsourced deliberation; deliberative democracy; participatory democracy

ACM Classification Keywords
H.5.3 [Information interfaces and presentation (e.g., HCI)]: Group and Organization Interfaces—Computer-supported cooperative work

Introduction
Crowdsourcing in policymaking has become a more common method deployed by local and national governments [1, 3, 17, 11]. Governments use crowdsourcing for knowledge search and civic engagement. They hope to find use-

ful knowledge for improving policies and to activate citizens to deliberate about policy issues on crowdsourcing platforms. Crowdsourced policymaking provides a space for democratic deliberation, in which the participant crowd exchanges arguments in the crowdsourcing process [2].

It remains unknown, however, what the quality of deliberation in crowdsourced policymaking is. Crowdsourced deliberation could, potentially, scale up to masses, unlike many other democratic innovations, such as citizen juries and deliberative polls, that serve only a small-scale participation [19]. But if the quality of crowdsourced deliberation is low, scaling up is not meaningful. To address this question, we examine the deliberative quality in two crowdsourced policymaking processes by using the Discourse Quality Index (DQI). The goal is to provide empirical insights that inform debates about the value of crowdsourced policymaking as a democratic innovation and a method for participatory democracy, while also offering insights into the crowd's behavior in crowdsourced policymaking.

Theoretical Background

Crowdsourcing in Policymaking

When crowdsourcing is deployed in policymaking, the online crowds are invited to contribute to a policy with their ideas. The crowd can attend several parts of the policymaking process, such as developing options and evaluation. The crowd is, however, rarely a part of decision-making. In crowdsourced policymaking orchestrated by governments, it is the elected decision-making body that decides about the policies, while the crowd serves as an additional data point to legal drafters. Thus, crowdsourced policymaking is a method for participatory democracy [16], not for direct democracy, because the crowd does not have decision-making power.

Crowdsourcing is applied in several ways in policymaking: as crowdsourced microtasking, ideation, argumentation, and deliberation [2]. The crowdsourced knowledge can be ideas or situated knowledge, which can help the policymakers to formulate a stronger policy. In crowdsourced argumentation and deliberation, the crowd is asked to exchange arguments about a given topic, as on dedicated deliberation platforms such as Deliberatorium [9], Consider.it [10], and Regulation Room [18].

Democratic Deliberation

Democratic deliberation is "the public use of arguments and reasoning among free and equal individuals" (adapted from [6], c.f. [13]). Deliberation requires a reasoned exchange of arguments, and democratic deliberation requires equal standing among free participants ("free and equal") and a public, to a certain degree transparent exchange. The core features of democratic deliberation are the presence of arguments and critical listening among free and equal participants [2]. Deliberative democrats advocate for democratic deliberation for its epistemic and legitimacy-enhancing capacities: It is argued to create a more informed and active citizenry, awareness of societal issues and learning, while the participatory nature of the process enhances the legitimacy of the outcome [14].

The golden standard for democratic deliberation is set in the minipublics approach [12], aiming to detect the public opinion by gathering a group of citizens to deliberate about a given issue, as in deliberative polling [7] and citizen juries. The participants are recruited by random sampling. While providing a good setting for deliberation, the challenge of minipublics is their small scale and high cost of in-person deliberations [20].

Crowdsourced Deliberation

Crowdsourced deliberation in policymaking, instead, can scale up to masses. It combines the core characteristics of democratic deliberation and crowdsourcing [2]. Crowdsourced democratic deliberation is an open, asynchronous, depersonalized, and distributed online deliberation occurring among self-selected participants, initiated by government or another organization orchestrating the crowdsourced policymaking. It includes reasoned argumentation, equality, and publicity, and the process is controlled by the crowdsourcer [2]. It is a part of a larger policy-reform process, governed by the crowdsourcers to develop a stronger policy, thus differing from online deliberation on other online fora such as newspapers' commenting sections or Reddit comment threads.

Evaluating the Quality of Deliberation

Discourse Quality Index (DQI) is the most used method in analyzing deliberative quality [4, 20], rooted in the Habermasian discourse ethics [8]. In DQI, the participants' comments are analyzed by the following categories: equality; justification rationality; common good orientation; respect, and constructive politics [21]. These have threshold values for quality in deliberation [4]. In addition, the nature of deliberation is analyzed by the following categories: interactivity/reciprocity, constructive character of the exchanges, deliberative negotiations, storytelling, and type of justification [4, 15].

Prior work with evaluation of deliberative quality with DQI has found that the quality of deliberation varies based on topics. The ideal, perfect deliberation remains beyond reach [4, 15, 20]. Prior research has focused on analyzing deliberative quality mostly in recruited minipublics and in offline deliberations, and there is a lack of knowledge about the deliberative quality of deliberation in crowdsourced policymaking.

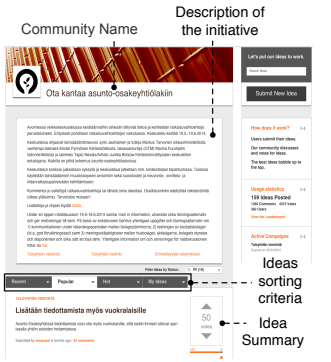
Therefore we address the following questions: How is the quality of deliberation in crowdsourced policymaking? What are the characteristics of crowdsourced democratic deliberation in policymaking? How does the deliberative quality vary among participants in crowdsourced policymaking? Addressing these questions is important for identifying the deliberative capacities of crowdsourced policymaking, so that they can be enhanced with appropriate process and technology designs.

Case Profiles, Data and Methods

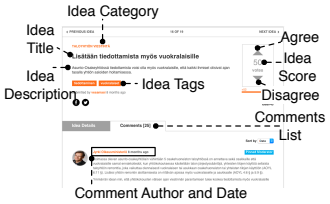
Case Profiles

We draw on data from two crowdsourced policymaking processes, an off-road traffic law reform (OTL) and a housing company reform (HL), both of which took place in Finland. The Ministry of the Environment in Finland deployed crowdsourcing in the off-road traffic law reform in two sequences in the spring of 2013. The crowd was asked to submit ideas for improving the law in categories, including safety, age limits, protecting nature, and regulation of the route establishment process. The prompts for the crowd included information about the law and questions for them to answer. The crowdsourcing resulted into 500 ideas and 4000 comments from over 700 users.

In the second case, the Finnish Ministry of Justice applied crowdsourcing in the reform of the Limited Liability Housing Companies Act (HL). The law regulates the governance, management, and administration practices of most condominium (apartment) buildings in Finland. The first sequence of crowdsourcing began in May 2014 and ended in June. In June 2015, the ministry launched the second crowdsourc-



(a) Home page on the crowdsourcing platform



(b) Idea View

Figure 1: (a) Crowdsourcing platform UI; (b) Detailed view of an idea, commenting and voting functions

ing stage. 556 participants registered on the platform, producing 232 ideas and 2,901 comments.

In both processes, crowdsourcing took place on an online platform, running on customized commercial crowdsourcing software. The participants could propose ideas on the platform, comment, and like or dislike ideas by using a thumbs-up/thumbs-down modality, as Figure 1 illustrates. The process and the platform was designed for knowledge search, and the crowd-input was visible to the online public. To participate, the users had to register on the site with a verifiable email. They could choose to stay anonymous or use their real names. The process was post-moderated.

Data and Methods

We used Discourse Quality Index (DQI) for analyzing interactions in the crowdsourced policymaking processes. Two coders coded participants' comments by using the DQI categories, as Table 1 shows. In addition, the type of justification was coded by using the categories introduced by Boltanski and Thevenot [5] in their theory of pragmatic justification (civic, legal, ecological, domestic, aesthetic, industrial, proximity, health, solidarity, safety, efficiency), added by categories that emerged in the data: transparency, and harm: smell and noise, as depicted in Table 2. We added a category for the nature of the tone and content sharing: links, pictures, and videos. See Table 3 in the supplementary materials for more details about the categories.

Two coders coded 1044 ideas and comments (74% of the total) in HL, and 894 ideas and comments (29% of the total) in OTL. The coders reached 73% inter-coder reliability in HL and 74% in OTL. The DQI scores were calculated by using the procedure by Bachtiger et al. [4] and Steenbergen et al. [21]. In addition, we used the activity logs (number of ideas and comments) on the online platform to analyze the users' activity. We used surveys to gather demographic information from the participants. In OTL, out of 743 registered users, 186 replied, resulting to a 25% response rate. Of the 556 participants in HL, 316 people took the survey, with a 57% response rate.

Findings

Reciprocity, respect and story-telling

The deliberation in the crowdsourced policymaking processes had a high level of implicit respect for other groups and their demands. There was a high level of reciprocity (interactivity), meaning, the participants referred to each others' comments, indicating they had read them. The deliberation had a considerable amount of storytelling, and there were some deliberative negotiations, as Table 1 shows. In most interactions, there was inferior justification where the linkage between reasons and conclusion is tenuous, and more than one-third of the interactions (HL 39%, OTL 35%) had at least qualified justification rationality with a linkage between reasons and conclusion. See the supplementary materials for an extended version of the Table 1. As Table 2 shows, the most commonly used justifications were efficiency, industrial (financial reasons) and legal reasoning, which are common-sensical considering the areas of the crowdsourced laws. There was little common-good orientation in the justifications. Most interactions had either explicit statement concerning group interests or neutral statements instead of statements for the common good in utilitarian terms or for helping the least advantaged in the society.

The overall tone of the interactions was neutral. There were no flame wars, blaming, or shaming. In five (equality, common-good orientation, respect groups, respect demands others, and constructive politics) of the seven indicators, that have threshold values for the quality of democratic deliberation set by [4], the criteria are met, showing that the deliberative quality meets the standards set by de-

Case	*Indicator*	*Mean*	*Median*	*SD*	*Min*	*Max*	*Number of cases per value (percentage of total)*				
	Participation Equality [0-1]						0	1			
HL		1	1	0	1	1	0	1044 (100%)			
OTL		1	1	0	1	1	0	894 (100%)			
	Level of Justification Rationality [0-4]						0	1	2	3	4
HL		1.1	1	0.99	0	3	372 (36%)	265 (25%)	318 (30%)	89% (9%)	0
OTL		1.1	1	0.95	0	4	265 (30%)	316 (35%)	232 (25.9%)	80 (9%)	1 (0.1%)
	Common-good Orientation [0,1, 2a, 2b] *						0	1	2a	2b	
HL		0.49	0	0.52	0	2	544 (52%)	491 (47%)	6 (0.6%)	3 (0.4%)	
OTL		0.49	0.5	0.50	0	2	477 (53%)	391 (44%)	22 (2.5%)	4 (0.5%)	
	Respect Groups [0-2]						0	1	2		
HL		0.88	1	0.37	0	2	143 (14%)	877 (85%)	14 (1%)		
OTL		0.96	1	0.23	0	2	42 (5%)	845 (94%)	7 (1%)		
	Respect Demands Others [0-2]						0	1	2		
HL		1.03	1	0.38	0	2	61 (6%)	888 (85%)	95 (9%)		
OTL		0.91	1	0.45	0	2	133 (15%)	708 (79%)	53 (6%)		
	Respect Counterarguments [0-3]						0	1	2	3	
HL		0.24	0	0.66	0	3	904 (86%)	41 (4%)	82 (8%)	17 (2%)	
OTL		0.36	0	0.66	0	3	650 (73%)	184 (20%)	43 (10%)	17 (2%)	
	Constructive Politics [0-2]						0	1	2		
HL		0.1	0	0.41	0	2	984 (94%)	17 (2%)	43 (4%)		
OTL		0.2	0	0.55	0	2	787 (88%)	40 (4%)	67 (8%)		
	Reciprocity [0-1]						0	1			
HL		0.9	1	0.31	0	1	111 (11%)	933 (89%)			
OTL		0.9	1	0.34	0	1	120 (14%)	774 (86%)			
	Story-telling [0-1]						0	1			
HL		0.35	0	0.48	0	1	682 (65%)	362 (35%)			
OTL		0.15	0	0.36	0	1	760 (85%)	134 (15%)			
	Deliberative Threats [0-1]						0	1			
HL		0.002	0	0.04	0	1	1042 (99.8%)	2 (0.2%)			
OTL		0.04	0	0.19	0	1	859 (96%)	35 (4%)			
	Deliberative Promises [0-1]						0	1			
HL		0.01	0	0.10	0	1	1033 (99%)	11 (1%)			
OTL		0.1	0	0.26	0	1	831 (93%)	63 (7%)			
	The Tone [0-2]						0	1	2		
HL		1.24	2	0.84	0	2	271 (26%)	294 (24%)	524 (50%)		
OTL		1.45	2	0.81	0	2	180 (20%)	128 (14%)	586 (76%)		

Table 1: DQI results in crowdsourced off-road traffic law (OTL) and housing company law (HL). The bolded indicators meet the level for quality in deliberation set by [4]. The indicators with a statistically significant difference between the cases are italicized ($p - value < 0.05$). * For the calculation of the descriptive statistics of this indicator 2a and 2b were represented as 2.

	HL	OTL
Efficiency	14%	17%
Industrial	15%	15%
Legal	17%	13%
Domestic	1%	15%
Solidarity	7%	6%
Noise	0.1%	11%
Civic	5%	3%
Proximity	1%	3%
Ecological	0.1%	23%
Aesthetic	0.1%	2%
Smell	0%	2%

Table 2: Types of justification rationality ordered from highest to lowest vales considering the mean of the two values

liberative democrats. The other descriptors of deliberative quality (reciprocity, storytelling, deliberative negotiations: threats/promises) had considerably higher occurrences than in prior research [4, 15]. In sum, there was a reasoned exchange of arguments between free and equals, respect, listening and reciprocity. The main difference between the cases was in storytelling: In HL, 35% of the comments had storytelling, whereas in OTL, it was 15%. This reflects the different natures of deliberation in the cases: In HL, the crowd shared stories of incidents in apartment buildings to illustrate their arguments. In OTL, the discussion was more abstract.

The measures of deliberative quality that scored clearly lower were justification rationality and respect towards counterarguments. And, of course, all the scores could be better—the Weberian ideal type of Habermasian democratic deliberation was not reached, yet it is not reached in prior work either. DQI was initially developed for in-person, synchronous deliberations in political contexts, thus emphasizing the common-good orientation and justification rationality, which experienced political actors can be expected to master. DQI has been adjusted to citizen deliberation [4], and the other characteristics of democratic deliberation, such as reciprocity, deliberative negotiations, and storytelling in DQI, are essential in characterizing crowdsourced deliberation.

Super-deliberators vs. low-quality deliberators

We identified a group of participants, whose interactions scored consistently higher than others. These "super deliberators" were mostly men in OTL and women in HL, and their participation activity varied. They had a mixed educational background, and varied civic activity levels. There were two distinct patterns in their participation behavior: In OTL, they were typically the last ones to participate in popular comment threads, indicating that super-deliberators join discussions after ideas become popular and/or controversial. Whereas in HL, they were usually the first ones commenting on the threads. This deserves further investigation to detect the crowd's behavioral patterns in more detail. There is another group, whose interactions scored consistently lower than others. In HL, the low-quality deliberators were mainly men of 65 years old or more. They had a mixed educational background and were not particularly civically active, but participated on online forums. In OTL, most of the low-quality deliberators were 26-54 years old men with a mixed educational background. They were also not particularly civically active, but participated on online forums.

Discussion and Conclusions

The findings show that even without a design that would specifically support deliberative quality, deliberation in crowdsourced policymaking can meet the standards of good democratic deliberation. If the crowdsourcing platform was designed to enhance deliberative quality, the quality would most likely improve. The technology could, for example, encourage the user to provide more justifications for their arguments and prompt for storytelling for real-world examples in narratives.

The findings show the promise of crowdsourced deliberation as a valuable part of deliberative systems in the society. This challenges the belief that post-moderated, openly accessible crowdsourced deliberations allowing anonymous participation only attract trolls and low-quality interactions. Vice versa, the findings show that crowdsourced policymaking can provide value in creating spaces for democratic deliberation, holding the promise of scaling it up.

REFERENCES

1. Tanja Aitamurto. 2012. *Crowdsourcing for Democracy: A New Era in Policy-Making*. Publications of the Committee for the Future, Parliament of Finland. Helsinki, Finland.
2. Tanja Aitamurto and Hélène Landemore. 2016. Crowdsourced Deliberation: The Case of the Law on Off-Road Traffic in Finland. *Policy & Internet* 8, 2 (2016), 174–196.
3. Tanja Aitamurto, Hélène Landemore, and Jorge Saldivar. 2016. Unmasking the Crowd: Participants' Motivation Factors, Expectations, and Profile in a Crowdsourced Law Reform. *Information, Communication & Society* (2016), 1–22.
4. André Bächtiger, Susumu Shikano, Seraina Pedrini, and Mirjam Ryser. 2009. Measuring deliberation 2.0: standards, discourse types, and sequenzialization. In *ECPR General Conference, Potsdam*. 5–12.
5. Luc Boltanski and Laurent Thévenot. 2006. *On justification: Economies of worth*. Princeton University Press.
6. Joshua Cohen. 1989. *Deliberation and Democratic Legitimacy*. In The Good Polity, eds. A. Hamlin and P. Pettit. New York: Basil Blackwell. 17–34 pages.
7. James S Fishkin. 2011. *When the people speak: Deliberative democracy and public consultation*. Oxford University Press.
8. Jürgen Habermas. 1984. *The theory of communicative action, volume I*.
9. Mark Klein. 2011. The MIT deliberatorium: Enabling large-scale deliberation about complex systemic problems. In *Collaboration Technologies and Systems (CTS), 2011 International Conference on*. IEEE, 161–161.
10. Travis Kriplean, Jonathan Morgan, Deen Freelon, Alan Borning, and Lance Bennett. 2012. Supporting reflective public thought with considerit. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 265–274.
11. Hélène Landemore. 2015. Inclusive Constitution-Making: The Icelandic Experiment. *Journal of Political Philosophy* 23, 2 (2015), 166–191.
12. Jane Mansbridge. 2010. Deliberative polling as the gold standard. *The Good Society* 19, 1 (2010), 55–62.
13. Jane Mansbridge, James Bohman, Simone Chambers, David Estlund, Andreas Føllesdal, Archon Fung, Cristina Lafont, Bernard Manin, and others. 2010. The Place of Self-Interest and the Role of Power in Deliberative Democracy. *Journal of Political Philosophy* 18, 1 (2010), 64–100.
14. José Luis Martí. 2006. The Epistemic Conception of Deliberative Democracy Defended Reasons, Rightness and Equal Political. *Deliberative democracy and its discontents* 27 (2006).
15. Laurence Monnoyer-Smith and Stéphanie Wojcik. 2012. Technology and the quality of public deliberation: a comparison between on and offline participation. *International Journal of Electronic Governance* 5, 1 (2012), 24–49.
16. Carole Pateman. 1970. *Participation and Democratic Theory*. Cambridge University Press.
17. John Prpic and Prashant Shukla. 2013. The Theory of Crowd Capital. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*. IEEE, 3505–3514.

18. Cynthia R. Farina, Dmitry Epstein, Josiah B. Heidt, and Mary J. Newhart. 2013. Regulation Room: Getting "more, better" civic participation in complex government policymaking. *Transforming Government: People, Process and Policy* 7, 4 (2013), 501–516.

19. Graham Smith. 2009. *Democratic Innovations: Designing Institutions for Citizen Participation*. Cambridge University Press.

20. Graham Smith, Corinne Wales, Peter John, Sarah Cotterill, Pat Sturgis, Stoker Gerry, and Hisako Nomura. 2009. Measuring the 'deliberative quality' of an online experimental mini-public: Methodology and early results. In *European Consortium for Political Research General Conference, in Potsdam, Germany, September*.

21. Marco R Steenbergen, André Bächtiger, Markus Spörndli, and Jürg Steiner. 2003. Measuring Political Deliberation: A Discourse Quality Index. *Comparative European Politics* 1, 1 (2003), 21–48.