When Life Gives You Lemons: How rating scales affect user activity and frustration in collaborative evaluation processes

Thomas Wagenknecht[1], Jan Crommelinck[1], Timm Teubner[2], and Christof Weinhardt[2]

[1] FZI Research Center for Information Technology, Berlin, Germany
{wagenknecht,crommelinck}@fzi.de
[2] Karlsruhe Institute of Technology, Karlsruhe, Germany
{timm.teubner,weinhardt}@kit.edu

Abstract. Initiators of open innovation processes involving customers or employees often face vast amounts of idea proposals. These proposals vary greatly in terms of quality, which is why organizers often engage the users themselves in the evaluation process. Building on the concept of information overload, we evaluate the effects of three distinct rating scales on users' activity and frustration measures. On the basis of an open innovation campaign for employees of a public-private institution in Germany, we systematically compare the novel "bag of lemons" method with conventional Likert scales and up-down-voting schemes. Our results demonstrate that the "bag of lemons"-approach yields higher levels of user activity, but is also perceived as significantly more frustrating. We find this effect to be fully mediated by perceived information overload, which points to potential avenues for the design of stimulating yet tolerably complex Information Systems for open innovation and rating techniques.

Keywords open innovation, rating scales, information overload, participation

1 Introduction

From strategic planning to product innovation, small and large firms as well as other organizations are involving their employees and stakeholders to propose novel ideas through digital platforms [1–3]. These processes are sometimes strictly limited to participation within the company or part of a larger open innovation campaign, including customers, suppliers, and other interested parties [4, 5]. Regardless of their target group, these platforms all have in common that users face vast amounts of proposals of varying quality, but only a few can or even should be implemented [2, 6]. Hence, there is a strong need for group decision support systems (GDSS) that enable users to filter ideas appropriately [7], i.e., that achieve high accuracy in identifying the best ideas and avoid to expose users to the adverse effects of information overload, including frustration and disengagement [8, 9].

13th International Conference on Wirtschaftsinformatik,
February 12-15, 2017, St. Gallen, Switzerland

Wagenknecht, T.; Crommelinck, J.; Teubner, T.; Weinhardt, C. (2017): When Life Gives You Lemons: How rating scales affect user activity and frustration in collaborative evaluation processes, in Leimeister, J.M.; Brenner, W. (Hrsg.): Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), St. Gallen, S. 380-394

Accordingly, there exists a myriad of filtering techniques. On the one hand, these include complex approaches such as prediction markets [10, 11], or automated methods like text mining that initially require a lot of human oversight and implementation capacity [12]. On the other hand, approaches like voting and user ratings are easier to implement and widespread on various online platforms. For instance, many social media and community platforms offer simple up- and down-voting (e.g., Reddit, Quora, Stackoverflow, etc.) or up-voting only (e.g., Facebook and Yammer in the form of "Likes"). Other platforms use Likert scales, often in form of star-ratings (e.g., Amazon, Airbnb, etc.). Yet, these methods face inherent shortcomings, including biased distributions [13], limited accuracy due to oversimplification, a possible disconnect between the goals of process organizers and raters, as well as reduced user satisfaction [14–16]. In this vein, the video platform YouTube dropped its Likert scale rating system in 2009 as users mostly rated content as either very bad or very good – rarely using any measures in the middle of the 5-point scale. Since then, the platform switched to up- and down-voting [17].

Seeking to address some of the shortcomings of existing approaches, Klein and Garcia [7] proposed a novel method. Their so-called "bag of lemons" (BOL) approach lets users in evaluation tasks allocate a predefined amount of *lemons* to those ideas they consider to be the worst. A lemon thus represents a negative assessment and a user can allocate multiple, indeed up to all of her lemons to one single idea. This way, the crowd is assumed to flag bad ideas, supposedly identifying a (remaining) set of high quality ideas. In fact, the BOL method outperformed Likert scales in terms of time for task completion and accuracy [7]. To follow up on these first auspicious insights, this paper systematically assesses the characteristics of the BOL method in comparison to up-/down voting and (conventional) Likert scales. In doing so, we focus on two factors. First, as crowd-based schemes rely on the laws of large numbers and the quality of collaborative evaluations usually increases in the number of independent assessments [18], we consider *user activity* under the three mentioned rating method regimes. Second, as crowd-based approaches typically work on a voluntary basis and hence require a positive user attitude and engagement [10, 19, 20], we consider the – potentially detrimental – effects on *frustration* as a key indicator of a non-positive attitude and user disengagement [20]. Such motivational variables are widely perceived as a crucial factor for user acceptance and usage of information systems [21, 22]. In this sense, this research is motivated by the following key drivers: First of all, there exists a clear research gap as BOL represents a novel method and its role in contrast to established methods is still unclear. However, organizations increasingly seek to involve their employees, citizens, or members in decision making in order to increase content, loyalty, identification, and productivity – often using those very collaborative voting techniques [23]. In consequence, as accruing informational charges grow constantly, such methods may expose participants to excessive informational load, yielding undesired results such as frustration, disaffection, and disengagement [8].

To connect the different rating methods with our target variables, we hence base our research on two intermediate, explanatory factors. First, as the BOL method represents a novel and commonly unknown rating technique, we consider the factor of *perceived novelty*, capturing potential user deterrence by the unknown, or a lack of

comprehensibility. Second, as BOL requires users to deal with a host of informational bits and pieces, *information overload* may be a concern. It was shown to yield adverse effects on employees as they are exposed to ever-growing amounts of unrestricted and unfiltered data [8, 9]. Accordingly, in this study, we pose the following overarching research questions:

RQ_1: How does the "Bag of Lemons" rating method affect user activity and frustration in a collaborative evaluation task?

RQ_2: Which role do perceived novelty and information overload play in mediating these effects?

To address our research questions, we conduct an online-based field experiment, including the collection of survey data. As part of a real world open innovation campaign, employees of a private-public institution rated the idea proposals of their peers. We systematically vary rating method, using up-/down voting, Likert scales [24] and the BOL method [7]. We investigate the ramifications for user activity, frustration [20], and task completion time, taking into account the factors perceived novelty and information overload [8]. Exceeding previous studies [7, 15, 20], users in this scenario were not forced to rate all ideas, which promises a more realistic situation and novel findings. In consequence, this study makes three main contributions to the Information Systems (IS) literature. First, we evaluate a novel, thus hardly researched method of idea evaluation (BOL) in comparison to more established methods (Likert scales, up-/down voting) in terms of the important indicators *user activity* and *frustration*, which has not or only scarcely been assessed by extant literature. By integrating these opposing factors within a joint research model, we enhance the understanding of collaborative evaluation processes in view of differentiated rating regimes [5, 25, 26]. Second, by relating these key indicators to mediating factors, we provide starting points for understanding *how* the different rating methods affect the users' perceptions and behaviors. In particular, we identify perceived information overload as a potential mediating factor at play. Third, our study provides a show case of employee-driven innovation [27] and computer-supported organizational participation [28].

This paper is organized as follows. We outline related work and the theoretical background in Section 2. Section 3 then illustrates our study design. Section 4 presents the results of our study. Lastly, we discuss our findings in view of theoretical and practical implications, limitations, and starting points for future research in Section 5.

2 Theoretical Background and Related Work

In recent years, the IS literature has begun to systematically evaluate ways to exploit the wisdom of the crowd, including a broad strand of research on open innovation processes [5, 29]. Notably, a number of studies analyzed voting and rating techniques on open innovation contests [7, 10, 15, 20, 30]. Such approaches relate to GDSS in the sense that groups evaluate proposals which were generated by the group itself, which can have important ramifications due to personal or social attachment, preoccupation, and other biases [31, 32]. With the emergence of large-scale open innovation contests,

IS research revived its investigation of rating scales. Several studies in this line of research evaluated both quality and task completion time with regard to different rating techniques [15, 20, 30, 33, 34]. In this section, we describe the theoretical background of the concepts and factors that form the basis of our study. We begin with a brief introduction of open innovation contests.

2.1 Open Innovation

Adamczyk et al. [5] define open innovation contests as IT-based and time-limited competitions by individuals or organizations calling on the general public or a specific target group to propose innovative solutions. Thereby the organizers make use of the expertise, skills, and creativity of distributed crowds. Engaging employees and customers in open innovation processes can have several benefits for the organizers, including increased loyalty, brand image, and success in recruitment [35]. For an open innovation contest to be successful, previous research identified a number of factors. Organizers, for instance, need to express a sense of urgency and establish a trusted environment [14, 36]. Moreover, users might be motivated by gaining access to the knowledge of experts as well as receiving appreciation for their input by peers and organizers of the process [37]. Furthermore, extant research also established that collaborative tools drive increase the quality of results in open innovation engagements [23].

Recently, several leading IT corporations engaged both their customers and employees in open innovation contests. For instance, IBM's "Innovation Jam" resulted in 46,000 product ideas proposed by 150,000 participants [1], while users in Dell's ongoing "IdeaStorm" have generated more than 20,000 suggestions for product improvements thus far [6]. Open innovation contests among employees of a company are one application of employee-driven innovation [27, 28]. In the broader context of computer-supported organizational participation, these contests can be a way to actively provide employees the means to be part of the decision-making processes of their workplace, which was found to be related to increased employee commitment and productivity [28].

Considering the vast amount of ideas, it becomes more likely that an open innovation contest will produce more superior solutions than an innovation process limited to only few innovators [38]. Thus, in line with the "wisdom of the crowds" paradigm, some user-generated ideas are able to compete with expert or core inside innovators [15, 25, 39]. However, assessing these crowd proposals can be costly. Robinson and Schroeder [40] estimate that large corporations take about four hours working time and $500 just to evaluate one idea. Yet, only few ideas are really worth increased attention. Prior research established that open innovation processes tend to produce large idea collections that are highly redundant and greatly vary in terms of quality [2, 20, 33, 39], where only 10-30% of the ideas tend to be of good or high quality [33]. Put figuratively, large-scale open innovation processes create excellent needles. They do, however, also create the corresponding haystacks. The main challenge then is to identify the valuable propositions. One common solution to this problem is to engage users in the evaluation process using voting and rating techniques [7, 10, 15, 20, 30, 34].

2.2 Rating Scales, Attitudes, and Intrinsic Motivation

The usage of rating scales transforms the process of idea evaluation into a concrete task of judgment, where individuals consider a finite set of alternatives [10]. In effect, this enables the organizers of open innovation contests to reduce their costs for idea evaluation by basing decisions on aggregated user ratings.

However, the gathered data may depend on the specific rating scale. Prior research suggests that rating scales are prone to selection biases and other dysfunctionalities [7, 10, 15, 20, 30, 33, 34]. For instance, some researchers claim that rating scales often fail to properly distinguish between medium/good and excellent ideas [7, 34]. Moreover, there may occur discrepancies between the initiator's and the participants' goals and intentions. While initiators would like the participants to evaluate as many ideas as possible thoroughly, the latter are restricted both in terms of time and information available to them. Hence, organizers need to take potential factors such as non-interest, distractions, lack of knowledge, and workload into account [7, 20]. In consequence, they need to communicate clearly what, why, and how they would like their participants to do specifically.

Nonetheless, evaluation tasks are often described poorly and hence remain fuzzy. The rating scale itself hence become an important factor as participants are searching for potential cues [41]. In fact, participants tend to develop attitudes toward rating scales based on characteristics such as graphical elements and input variables [10, 19, 20]. Attitudes, in turn, can affect cognition and behavior [42]. In this context, Riedl et al. [20] found that users perceive different rating scales as more or less exciting, entertaining, satisfying, and positive, which can be explained by flow theory [43], suggesting that people can become very immersed by an activity, accompanied by high concentration on a task, while losing self-consciousness. Koufaris [44] suggested that flow states are related to increased intrinsic enjoyment and perceived control. Both constructs are also related to intrinsic motivation [45]. IS research established intrinsic motivation to be an important factor in creating favorable user perceptions, intention, and actual system use [21, 22]. In contrast, all too simple or overwhelmingly complex systems may deter users from entering such states, rendering system use a frustrating experience which is in consequence unlikely to be continued. Several potential antecedents of frustration come to mind. Given the structure of evaluation tasks with many diverse ideas, information overload is a concern which we further outline in the next paragraphs.

2.3 Information Overload

Information overload can be characterized as a state in which cognitive processing capacity is exceeded by the volume and speed of incoming stimuli that need to be processed [8]. People continuously evaluate their usage of information systems and discontinue usage when experiencing techno stress [46]. For instance, Maier et al. [47] found that users stop using social network services when experiencing, among other factors, exhausting levels of information disclosures by friends leading to information overload. Koroleva et al. [48] found similar results for Facebook and Eckhardt el. [49]

did so, asking participants in an experiment on LinkedIn to extract specific information for a job application. The phenomenon of information overload might be especially pronounced in open innovation evaluation tasks as users need to process a manifold, diverse, partly contradicting, and often novel set of ideas. Aggravatingly, the proposers usually do not follow a common schema, style, or language in describing their ideas. Comparing ideas across one another may hence be particular challenging.

Depending on the structure of the rating scale and evaluation task, perceived information load may thus differ [8]. It has, however, not been investigated with regard to rating scales in IS studies thus far. In the following, we hence describe a design allowing to relate users' perceptions of information overload to different rating methods, forming the basis of the field experiment reported in this paper.

3 Experimental Design

In this section, we outline an approach to address our research questions. Similar to Klein and Garcia [7], our study is based on an (internal) open innovation campaign at an actual private-public research center. Both the ideation as well as the evaluation phase were part of a broader participatory process at this institution [28]. The institution is legally incorporated as a foundation, disposes over a yearly budget of approximately €14 million, and employs a total of 280 people. Employees work on a variety of projects in the domains of computer science, information technology, robotics, and engineering.

Our study employs a two-staged approach. In the first stage, employees of this institution were invited to propose ideas on how to make the research center an (even better) employer via an online system. We invited all employees to this online platform. In the second stage, all employees were invited again to rate the ideas in a condensed set, using either BOL, up- and down-voting, or Likert scales.

3.1 Stage 1: Idea Generation

Employees of the institution were asked to propose ideas on how to make the center an (even better) employer. In a first phase, we received a total of 71 "raw" proposals. Before proceeding to the second stage, we eliminated hoax and proposals not compliant with the terms of use (e.g., including clear names of employees or foul language), consolidated redundant proposals, redacted grammatical and other language- and style-related issues, and in consequence, generated a condensed and workable idea corpus of 42 proposals. The proposals covered a wide range of topics, addressing organizational procedures, marketing, human resources, and many other areas. In this first stage, participants were able to propose ideas within a range of two weeks. Ideas were generally posted anonymously in order to both comply with German data protection legislation and to enable employees to speak their mind freely [50, 51].

3.2 Stage 2: Idea Evaluation

In the second stage, employees were then invited to rate their peers' proposals on another online platform. This platform was accessible for two weeks, too. Here, each

employee could participate only once. Participants were prompted to assess the ideas' overall quality, which may be based on subcategories such novelty, feasibility, or value to the company [15, 20, 30, 39]. Note, however, that these sub-dimensions were not surveyed separately. In fact, idea evaluation was based on either bag of lemons, up- and down voting, or Likert scales.

Each participant was allocated to only one of the three treatment conditions (between-subjects design). All participants were presented the same 42 proposals in all treatment conditions, using a random order for each participant in order to rule out sequence effects. Following Klein and Garcia [7], participants in the BOL setting disposed over a total of eight lemons, representing ~20% of the total idea basket, which they were able to allocate to the ideas. In the up- and down-voting setting, participants could either up-vote or down-vote each idea once. This setting replicates that of platforms such as YouTube. Participants in the Likert scale setting were able to rate the ideas on 5-point Likert scales, ranging from 1 (very bad) to 5 (very good). Exceeding previous studies [7, 15, 20], participants in the Likert and up- and down-voting treatments were free to rate as many ideas as they liked, that is, there was neither a minimum nor maximum requirement. Participants were asked to complete a mandatory quiz before the actual rating task in order to ensure comprehension and hence validity.

3.3 Measures

After completing the rating process, participants were asked to conduct a brief survey. To ensure validity, previously validated scales were used and adapted to the context of this study. We assessed user attitudes towards the rating method, operationalized by the categories novelty and frustration [20, 52]. Information overload was adopted based on the items proposed by Schultz and Vandenbosch [8]. To assess user activity, we measured how many votes were casted in relation to the maximum number of votes in the respective treatment. This index ranges between 0 and 1.

Table 1. Measurement items

Construct	***Item***	***Source***
Perceived Novelty	Using the rating scale was a novel experience to me.	[20]
Frustration	Using the rating scale was a frustrating experience to me.	[20]
Information Overload	In using the rating scale, I was forced to concern myself my many idea proposals.	[8]
	In using the rating scale, I could not focus on the actual relevant idea proposals.	
	The rating scale overcharged me by too many idea proposals and too much information.	

4 Results

In total, 141 participants completed the questionnaire, representing approximately 50% of the total workforce at the institution. Altogether, 54 participants evaluated the ideas using BOL, 48 were in the Likert treatment, and 39 in the up- and down-voting

treatment. In compliance with German privacy regulation, participants were able provide personal information on a voluntary basis. Thus, only part of our sample reported age (61.5%) and/or gender (71.5%). The age of the (reporting) participants ranged from 18 to 37 years (mean 28.9). Moreover, 80% of our participants were male. These characteristics did not differ significantly among the three treatments.

We first turn to the central target measures of this study, user activity and frustration. As illustrated in Figure 1, user activity was highest for the BOL method, and lowest for up-/down voting. A set of t-test confirms the significance of these differences ($t_{BOL/Likert}$ = 1.648, p=.103; $t_{BOL/U\&D}$ = 4.347, p<.001; $t_{Likert/U\&D}$ = 3.206, p<.001). As a first result, we thus note that the bag of lemons rating scheme facilitates higher levels of user activity than Likert scales or up- and down voting.

Next, we consider how frustrating users perceived the different rating methods. Figure 1 shows that BOL provokes markedly higher levels of frustration than the other methods, whereas Likert and up-/down voting yield comparable levels. A set of t-test confirms this impression statistically ($t_{BOL/Likert}$ = 2.498, p=.014; $t_{BOL/U\&D}$ = 2.783, p=.007; $t_{Likert/U\&D}$ = .283, p=.778). As a second result, we note that the bag of lemons rating scheme facilitates higher levels of perceived frustration than Likert scales or up- and down voting.

Besides these focal measures, we surveyed the participants in terms of how novel and how (informational) overloading they perceived the three rating methods. As can be seen in Figure 1, both for novelty and information overload, the bag of lemons method yields (marginally) significant higher levels than the other two (Novelty: $t_{BOL/Likert}$ = 11.033, p<.001; $t_{BOL/U\&D}$ = 11.711, p<.001; $t_{Likert/U\&D}$ = .983, p=.328; Overload: $t_{BOL/Likert}$ = 1.816, p=.072; $t_{BOL/U\&D}$ = 2.555, p=.013; $t_{Likert/U\&D}$ = 1.0613, p=.292).

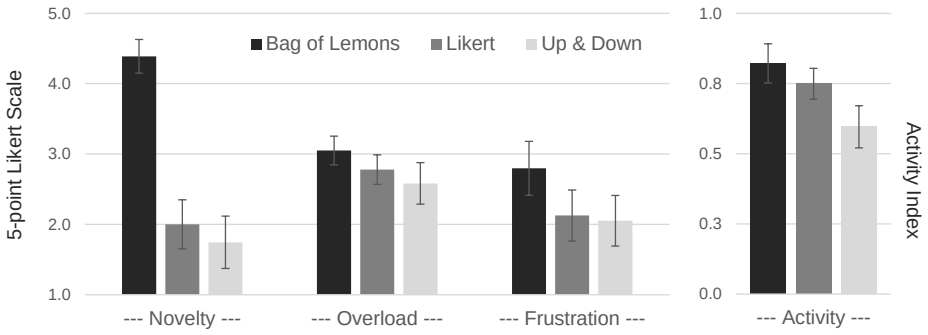

Figure 1. Overview of novelty, information overload, frustration, and activity scores (error bars indicate 95% confidence intervals)

We now turn to a structural analysis of the effects of rating scale on user activity and frustration. As we have outlined in Section 2, we hypothesize perceived novelty and information overload as potential mediators, that is, carriers and hence psychological determinants of the rating scale effects on the target measures. For doing so, we slightly simplify the analysis, comparing the bag of lemons method against both

other methods simultaneously, that is, using only one binary dummy variable for "bag of lemons." Our model, along with the results, is depicted in Figure 2. We use structural equation modelling based on partial least squares (SEM-PLS) to operationalize this analysis. Specifically, SmartPLS 3.0 [53] was used due to its flexibility in terms of sample size and its lack of assumptions regarding data and residuals distribution [54]. The sample size of this study (n = 141) exceeded the minimum required to validate a model in PLS, given the present structural model [55]. Confirming the results from above, this analysis shows that the bag of lemons significantly increases the perception both of (rating scale) novelty (b=.743, p<.001) as well as information overload (b=.212, p<.010). Information overload, in turn, significantly drives frustration (b=.262, p<.010), whereas the direct path from BOL to frustration is insignificant. Thus, information overload fully mediates the method's direct impact on frustration (beyond its indirect effect via this path).

In contrast, there does not occur any mediation on user activity, neither via perceived novelty, nor via information overload – both paths are insignificant. There exists, however, a positive and significant direct effect from BOL to user activity (b=.390, p<.001).

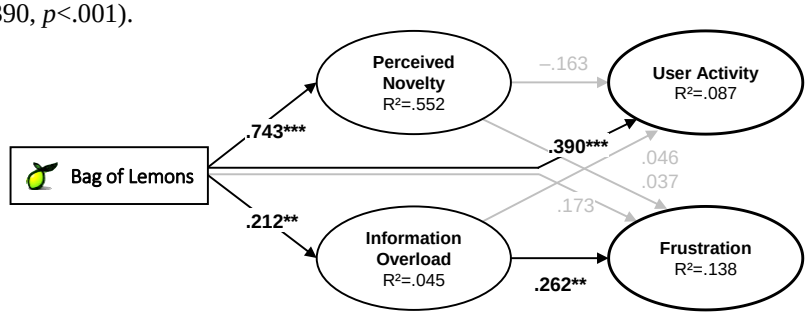

Figure 2. Structural Research model, including standardized path coefficients and R squared values (*** p<.001; ** p<.01)

Lastly, we considered the individual task completion times. Since this factor has an open-ended scale in one direction, Figure 3 depicts the main characteristics of the time distributions for the three treatment conditions in boxplot diagrams (indicating, median, as well as 25%- and 75%-quartiles). We find that the three conditions do not differ significantly in terms of completion time ($t_{BOL/Likert}$ = 1.564, p=.122; $t_{BOL/U\&D}$ = 1.467, p<.147; $t_{Likert/U\&D}$ = –.097, p=.923).

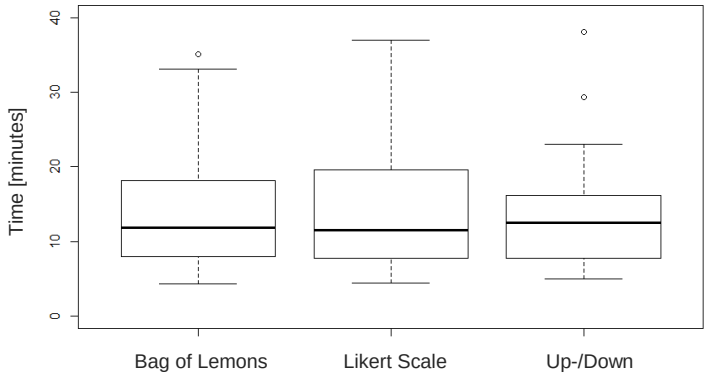

Figure 3. Boxplots of task completion times

5 Discussion and Conclusion

In this paper, we analyzed the effects of rating scales on users' activity, perceived information overload, perceived novelty, and frustration. In a field experiment in an open innovation campaign for a mid-size German research center, we assessed how BOL, up- and down-voting, and Likert scales differed in terms of these measures when employees were asked to evaluate a corpus of ideas created by their peers. All employees of the research center were invited to rate 42 proposals, being exposed to one of the above mentioned rating scales (between subjects design). Analyzing the behavioral as well as the post-evaluation survey data, we demonstrate that BOL, while stimulating activity, is also perceived as more frustrating than other rating techniques. We trace this result to the mediating factor of perceived information overload. Although participants were exposed to the same amount of information, that is, the identical corpus of 42 ideas, the bag of lemons method yielded much higher overload perceptions. We suggest that this may be due to deliberative and "pending" nature of the bag of lemons approach. While using Likert scales or up/down voting techniques, each idea can be assessed at a time, allocating lemons to a set of many ideas can be challenging since the desire to allocate a lemon late in the process may require to reassess previously rated ideas, for instance, to decide where to withdraw lemons from. This need for continuous cross-links requires to keep more ideas in mental "working memory," whereas they can be considered (and forgotten) sequentially when using the other techniques.

Coming back to our first research question of how the "Bag of Lemons" rating method affect user activity and frustration in a collaborative evaluation task, we hence can summarize that BOL increases both user activity and frustration. With regard to the second research question, that is, the role of perceived novelty and information overload in mediating these effects, we see that information overload fully mediates the effect of the BOL method on frustration, while perceived novelty does not exhibit any mediating properties. Moreover, there do not occur any cross-mediating effects, that is, from novelty to frustration or from information overload to activity.

Considering that approximately 50 percent of the employees of the institution evaluated their peers' proposals, this also hints at the high interest of employees in getting engaged in the process of participating in the decision-making processes at their workplace [28].

Theoretical and Practical Implications

This study contributes to the literature by evaluating a novel, thus hardly researched method of idea evaluation (BOL) in comparison to more established methods (Likert scales, up-/down voting). We focus on the important indicators of *user activity* and *frustration*, which has not or only scarcely been assessed by extant literature in this context. By integrating these opposing factors within a joint research model, we enhance the understanding of collaborative evaluation processes in view of differentiated rating regimes [5, 25, 26]. Next, by relating these key indicators to mediating factors, we provide starting points for understanding *how* the different rating methods affect the users' perceptions and behaviors. In particular, we identify perceived information overload as a potential mediating factor at play. Moreover, our study provides a show case of employee-driven innovation [27] and computer-supported organizational participation [28]. We confirm findings of Riedl et al. [20], who suggested that people form attitudes towards rating scales. Our findings also lend support to Klein and Garcia [7], underpinning BOL's novelty but, in contrast, do not confirm the method's superiority in terms of task completion time. Yet, we extend the authors findings by shedding light on users' perception of BOL's restraining character. Participants in our study expressed higher levels of frustration when evaluating ideas using the BOL as compared to Likert and up- and down-voting. This suggests that people might refrain from engaging in a BOL evaluation task in the future. Accordingly, practitioners should be aware of the possibly detrimental effects of BOL when designing an open innovation platform. This effect, as it is mediated by perceived information overload, may substantially be driven by the relatively high number of idea. We suggest that idea evaluation tasks with fewer ideas (e.g., 6 to 12), may yield different results.

Limitations and Future Research

Our study needs to be considered against several limitations. First, we compared the different rating methods in terms of user activity, frustration, and time, however, could not consider the evaluations' accuracy, that is, a match between the crowd's assessment versus how good the ideas actually were. This limitation points at several paths for future research, very much in the sense of prior studies [7, 20]. Future work needs to take into account accuracy, for instance by comparing the collaborative results with an expert rater panel.

Next, as we have shown in this study, BOL facilitates higher levels of (relative) user activity than other rating methods. Nonetheless, on average, Likert and up-/down votes yield a higher overall numbers of idea evaluations. Systematically varying the amounts of ideas and "lemons" to distribute could thus shed more light on the strengths and weaknesses of the BOL approach and its robustness against different set sizes.

Due to strict German data protection legislation at the workplace, we were only able to capture some demographic characteristics of our participants. Thus, the data set is

somewhat incomplete and restricts us from fully taking into account potential age or gender effects. Based on the data we have, these characteristics did not differ between treatments, so that at least a treatment bias due to demographic factors could be ruled out. Another limitation relates to the fact that part of the correlation between the item-based measures may be due to common method bias as most data was collected using standard questionnaire items. User activity represents an exception; correlations here will not exhibit common method bias.

As this study finds rating scales to affect user frustration, we suggest that it is worth exploring the antecedents of scale-related techno-stress. The noteworthy differences for information overload between BOL and up- and down-voting already lend some support to this presumption.

Furthermore, our study as well as previous ones [7, 10, 15, 20, 34] asked participants to rate ideas in the absence of any indication on whether and how other users already rated proposals. Future research could thus investigate the impact of information cascades, that is, users being able to see the evaluations of other (earlier) users [56], which may significantly impact results [57]. Finally, future research should address how open innovation contests within companies shape employee commitment and overall productivity [28].

6 Acknowledgement

This study was part of the joint research project “Participation as a Service” (PaaS), funded by the German Federal Ministry of Education and Research.

References

1. Bjelland, O.M., Wood, R.C.: An Inside View of IBM’s “Innovation Jam.” MIT Sloan Manag. Rev. 50, 32–40 (2008).
2. Di Gangi, P.M., Wasko, M.: Steal my idea! Organizational adoption of user innovations from a user innovation community: A case study of Dell IdeaStorm. Decis. Support Syst. 48, 303–312 (2009).
3. Niemeyer, C., Wagenknecht, T., Teubner, T., Weinhardt, C.: Participatory Crowdfunding: An approach towards engaging employees and citizens in institutional budgeting decisions. In: 49th Hawaii International Conference on System Sciences (HICSS). pp. 2800–2808 (2016).
4. Chesbrough, H.W.: Open Innovation: The New Imperative for Creating and Profiting from Technology. (2003).
5. Adamczyk, S., Bullinger, A.C., Möslein, K.M.: Innovation Contests: A Review, Classification and Outlook. Creat. Innov. Manag. 21, 335–360 (2012).
6. Hossain, M., Islam, K.M.Z.: Ideation through Online Open Innovation Platform: Dell IdeaStorm. J. Knowl. Econ. 6, 611–624 (2015).
7. Klein, M., Garcia, A.C.B.: High-speed idea filtering with the bag of lemons. Decis. Support Syst. 78, 39–50 (2015).
8. Schultz, U., Vandenbosch, B.: Information Overload in a Groupware

Environment : Now You See It, Now You Don't. J. Organ. Comput. Electron. Commer. 8, 127–148 (1998).
9. Oldroyd, J., Morris, S.: Catching Falling Stars: A Human Resource Response to Social Capital's Detrimental Effect of Information Overload on Valuable and Visible Employees. Acad. Manag. Rev. 37, 396–418 (2012).
10. Blohm, I., Riedl, C., Füller, J., Leimeister, J.M.: Rate or Trade? Identifying Winning Ideas in Open Idea Sourcing. Inf. Syst. Res. 27, 27–48 (2016).
11. Teschner, F., Rothschild, D.: Simplifying market access: A new confidence-based interface. J. Predict. Mark. 6, 27–41 (2013).
12. Martinez-Torres, M.R.: Content analysis of open innovation communities using latent semantic indexing. Technol. Anal. Strateg. Manag. 27, 859–875 (2015).
13. Teubner, T., Saade, N., Hawlitschek, F., Weinhardt, C.: It's only pixels, badges, and stars: On the economic value of reputation on Airbnb. In: Australasian Conference on Information Systems 2016 (2016).
14. Ebner, W., Leimeister, J.M., Krcmar, H.: Community engineering for innovations: the ideas competition as a method to nurture a virtual community for innovations. R&d Manag. 39, 342–356 (2009).
15. Riedl, C., Blohm, I., Leimeister, J.M., Krcmar, H.: Rating Scales for Collective Intelligence in Innovation Communities: Why Quick and Easy Decision Making Does Not Get it Right. In: Proceedings of the 31st International Conference on Information Systems, St. Louis (2010).
16. Negahban, S., Sewoong, O., Shah, D.: Iterative Ranking from Pair-wise Comparisons. 1–9.
17. YouTube: Five Stars Dominate Ratings, https://youtube.googleblog.com/2009/09/five-stars-dominate-ratings.html.
18. Marion K. Poetz and Martin Schreier: The value of crowdsourcing: Can users really compete with professionals in generating new product ideas? J. Prod. Innov. Manag. 1–31 (2012).
19. Kamis, A., Koufaris, M., Stern, T.: Management Information Systems Research Center, University of Minnesota. MIS Q. 32, 159–177 (2008).
20. Riedl, C., Blohm, I., Leimeister, J.M., Krcmar, H.: The Effect of Rating Scales on Decision Quality and User Attitudes in Online Innovation Communities. Int. J. Electron. Commer. 17, 7–36 (2013).
21. Venkatesh, V.: Creation of Favorable User Perceptions: Exploring the Role of Intrinsic Motivation. MIS Q. 23, 239–260 (1999).
22. Hwang, Y., Yi, M.Y.: Predicting the Use of Web-Based Information Systems: Intrinsic Motivation and Self-Efficacy. In: Proceedings of the 8th Americas Conference onf Information Systems. pp. 1076–1081 (2002).
23. Blohm, I., Riedl, C., Leimeister, J.M., Krcmar, H.: Idea Evaluation Mechanisms for Collective Intelligence in Open Innovation Communities: Do Traders Outperform Raters? In: ICIS 2011 Proceedings. pp. 1–24 (2011).
24. Likert, R.: A Technique for the Measurement of Attitudes. Arch. Psychol. 22, 1401–1455 (1932).
25. Leimeister, J.M.: Collective Intelligence. Bus. Inf. Syst. Eng. 4, 245–248

(2010).
26. Straub, T., Gimpel, H., Teschner, F., Weinhardt, C.: How (not) to Incent Crowd Workers: Payment Schemes and Feedback in Crowdsourcing. Bus. Inf. Syst. Eng. 57, 167–179 (2015).
27. Gressgård, L.J., Amundsen, O., Merethe Aasen, T., Hansen, K.: Use of information and communication technology to support employee-driven innovation in organizations: A knowledge management perspective. J. Knowl. Manag. 18, 633–650 (2014).
28. Wagenknecht, T., Filpe, R., Weinhardt, C.: Towards a Research Framework of Computer-Supported Organizational Participation. In: Tambouris, E., Panagiotopoulos, P., Sæbø, Ø., Wimmer, M.A., Pardo, T.A., Charalabidis, Y., Soares, D.S., and Janowski, T. (eds.) Electronic Participation: 8th IFIP WG 8.5 International Conference, ePart 2016. pp. 17–28. Springer Publishing (2016).
29. Wagenknecht, T., Crommelinck, J., Teubner, T., Weinhardt, C.: Ideate. Collaborate. Repeat. A Research Agenda for Idea Generation, Collaboration and Evaluation in Open Innovation. In: 13th International Conference on Wirtschaftsinformatik (2017).
30. Dean, D.L., Hender, J.M., Rodgers, T.L., Santanen, E.L.: Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation. J. Assoc. Inf. Syst. 7, 646–699 (2006).
31. Sia, C.-L., Tan, B.C.Y., Wei, K.-K.: Group Polarization and Computer-Mediated Communication: Effects of Communication Cues, Social Presence, and Anonymity. Inf. Syst. Res. 13, 70–90 (2002).
32. Sassenberg, K., Postmes, T.: Cognitive and social processes in small groups : Effects of anonymity of the self and anonymity of the group on social influence. Br. J. Soc. Psychol. 41, 463–480 (2002).
33. Blohm, I., Bretschneider, U., Leimeister, J.M., Krcmar, H.: Does Collaboration among Participants Le ad to Better Ideas in IT-based Idea Competitions? An Empirical Investigation. In: Proceedings of the Annual Hawaii International Conference on System Sciences. pp. 1–10 (2010).
34. Bao, J., Sakamoto, Y., Nickerson, J. V.: Evaluating Design Solutions Using Crowds. In: Proceedings of the Americas Conference on Information Systems. p. Paper 446. , Detroit, Michigan (2011).
35. Fuchs, C., Schreier, M.: Customer empowerment in new product development. J. Prod. Innov. Manag. 28, 17–32 (2011).
36. Hawlitschek, F., Teubner, T., Weinhardt, C.: Trust in the Sharing Economy. Die Unternehmung. 70, 26–44 (2016).
37. Leimeister et al.: Leveraging Crowdsourcing: Activation-Supporting Components for IT-Based Ideas Competition. J. Manag. Inf. Syst. (2009).
38. Lakhani, K.R., Jeppesen, L.B.: Getting unusual suspects to solve R&D puzzles. Harv. Bus. Rev. 85, 30–32 (2007).
39. Poetz, M.K., Schreier, M.: The value of crowdsourcing: can users really compete with professionals in generating new product ideas? J. Prod. Innov. Manag. 29, 245–256 (2012).
40. Robinson, A.G., Schroeder, D.M.: Ideas are free: How the idea revolution is

liberating people and transforming organizations. Berrett-Koehler Publishers (2004).
41. Schwarz, N.: Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation. Lawrence Erlbaum, Hillsdale, NJ (1996).
42. Solomon, M., Bamossy, G., Askegaard, G., Hogg, M.K.: Consumer Behaviour: A European Perspective. Pearson FT Prentice Hall, Harlow, UK (2006).
43. Csikszentmihalyi, M.: Beyond Boredom and Anxiety. Jossey-Bass, San Francisco (1977).
44. Koufaris, M.: Applying the technology acceptance model and flow theory to online consumer behavior. Inf. Syst. Res. 13, 205–223 (2002).
45. Deci, E.L., Ryan, R.M.: Intrinsic Motivation Inventory, http://www.psych.rochester.edu/SDT/ measures/intrins.html.
46. Beaudry, A., Pinsonneault, A.: Understanding user responses to information technology: a coping model of user adaptation. MIS Q. 29, 493–524 (2005).
47. Maier, C., Laumer, S., Eckhardt, A., Weitzel, T.: Online Social Networks As a Source and Symbol of Stress: an Empirical Analysis. In: Proceedings of the 33rd International Conference on Information Systems, Orlando 2012 (2012).
48. Koroleva, K., Krasnova, H., Günther, O.: "STOP SPAMMING ME!" - Exploring Information Overload on Facebook. In: Proceedings of the 16th Americas Conference on Information Systems (AMCIS). p. Paper 447 (2010).
49. Eckhardt, A., Maier, C., Buettner, R.: The Influence of Pressure to Perform and Experience on Changing Perceptions and User Performance: A Multi-Method Experimental Analysis. Proc. 33rd Int. Conf. Inf. Syst. 1–12 (2012).
50. Haines, R., Hough, J., Cao, L., Haines, D.: Anonymity in Computer-Mediated Communication: More Contrarian Ideas with Less Influence. Gr. Decis. Mak. Negot. 23, 765–786 (2014).
51. Wagenknecht, T., Teubner, T., Weinhardt, C.: The Impact of Anonymity on Communication Persuasiveness in Online Participation. In: Proceedings of the Thirty Seventh International Conference on Information Systems (ICIS) (2016).
52. Galletta, D.F., Henry, R.M., Mccoy, S., Polak, P.: Web site delays: How tolerant are users? Inf. Syst. Res. 17, 20–37 (2002).
53. Ringle, C.M., Wende, S., Becker, J.-M.: SmartPLS 3. , Bönningstedt (2015).
54. Chin, W.W.: The partial least squares approach to structural equation modeling. Mod. Methods Bus. Res. 295, 295–336 (1998).
55. Gefen, D., Straub, D.W., Boudreau, M.-C.: Structural equation modeling and regression: Guidelines for research practice. Commun. Assoc. Inf. Syst. 4, 1–77 (2000).
56. Bikhchandani, S., Hirshleifer, D., Welch, I.: A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. J. Polit. Econ. 100, 992–1026 (1992).
57. Duan, W., Gu, B., Whinston, A.B.: Informational Cascades and Software Adoption on the Internet: An Empirical Investigation. MIS Q. 33, 23–48 (2009).