Consensus Building in Open Source User Interface Design Discussions

Roshanak Zilouchian Moghaddam, Brian P. Bailey, and Wai-Tat Fu
Department of Computer Science
University of Illinois
Urbana, IL 61801
{rzilouc2, bpbailey, wfu}@illinois.edu

ABSTRACT
We report results of a study which examines consensus building in user interface design discussions in open source software communities. Our methodology consisted of conducting interviews with designers and developers from the Drupal and Ubuntu communities (N=17) and analyzing a large corpus of interaction data collected from Drupal. The interviews captured user perspectives on the challenges of reaching consensus, techniques employed for building consensus, and the consequences of not reaching consensus. We analyzed the interaction data to determine how different elements of the content, process, and user relationships in the design discussions affect consensus. The main result from this analysis shows that design discussions having participants with more experience and prior interaction are more likely to reach consensus. Based on all of our results, we formulated design implications for promoting consensus in distributed discussions of user interface design issues.

Author Keywords
Consensus, design discussion, open source software

ACM Classification Keywords
H.5.3 [Group and Organization Interfaces]: Asynchronous interaction - Web-based interaction

General Terms
Design; Human Factors

INTRODUCTION

In open source software (OSS) communities, many design decisions that shape the product's user interface are made through distributed discussions [5]. For example, to initiate a community discussion of a usability issue, a member describes the issue and others join the discussion to propose and debate design alternatives. These discussions typically unfold via mailing lists or Web forums linked to the OSS project. It is imperative for discussion participants to reach consensus on a design proposal to show strong support for integrating its implementation into the product distribution and to enhance participants' feeling of valued contribution. By *consensus*, we mean that participants are willing to commit to a proposal despite any remaining objections [6].

Because consensus building is a common and critical task in user interface (UI) design discussions in OSS, it is important to understand how often consensus is (not) reached, what techniques are utilized to foster consensus, and which elements of a design discussion affect consensus, among many other interesting questions. However, these questions have not been directly addressed by prior work.

For example, one thread of prior research has analyzed elements of design discussions such as participation [16], argumentation [4, 20], and tracking of design proposals [32], but has not targeted consensus building. In a second thread, researchers have conducted lab experiments to test how different factors of group work such as size, task, and anonymity affect decision quality and reaching consensus [3, 17, 25]. However, in lab experiments, it is difficult to simulate how consensus unfolds in real world design discussions, especially those in a mature OSS community.

In this paper, we report results of a study which examines UI design discussions in OSS communities from the perspective of consensus building. Part of our methodology consisted of conducting semi-structured interviews with designers and developers from the Drupal and Ubuntu OSS communities (N=17). The interviews captured user perspectives on key challenges of reaching consensus in the design discussions, the techniques utilized for promoting consensus, and the consequences of not reaching consensus.

To complement the interviews, we analyzed a large corpus of interaction data collected from the Drupal community to test how different discussion elements relate to consensus. Our data set included UI design discussions that did and did not reach consensus. From the interviews and other sources, we derived three categories of potential factors; content, process, and user relationships and operationalized them into 23 metrics. The metrics were calculated for each discussion in our data set and entered in a binary (consensus or not) logistic regression. The main result from this analysis shows that discussions having participants with more experience and prior interaction history are more likely to reach consensus. Finally, based upon all of our

CHI'12, May 5–10, 2012, Austin, Texas, USA.

results, we offer design implications for promoting consensus in distributed discussions of UI design issues.

RELATED WORK

We describe the concept of consensus and why it is important for analyzing UI design discussions. We then show how our work complements prior studies of design discussions in OSS and group decision making.

Consensus and its Importance for Design Discussions

In group decision making, consensus refers to when all participants are willing to commit to a proposal despite the fact that objections may remain [6]. The process of building consensus requires a good faith effort to meet the diverse interests of all participants as the process is considered as important as the final outcome. For example, this is often accomplished by encouraging those with dissenting views to propose or adapt existing ideas to meet their own interests without supplanting the interests of others [27]. Consensus building is an increasingly utilized technique for group decision making in complex task domains (e.g. UI design) where no one person has all the required expertise or resources needed to solve a given problem [27].

In this paper, we study consensus building in the UI design discussions in one established OSS community. Consensus building is an appropriate and important lens for viewing these types of discussions. For example, since almost all community decisions relating to the product's interface are made through distributed discussions, the ability to reach consensus can have a large impact on the quality of the product and satisfaction with the decision making process.

Studies of Design Discussions in OSS

Researchers have often utilized the open nature of design discussions in OSS to study different elements of the user interface and software design process. However studies of consensus in these discussions are conspicuously missing.

For example, in our own prior work [32], we studied how users propose, track, and debate design alternatives during UI design discussions in OSS. Twidale and Nichols [28] explored how usability issues are reported, discussed, and resolved in several OSS bug repositories with a goal of understanding how to improve the discussion interface. Ko and Chilana [20] also analyzed discussions in OSS repositories but with the goal of understanding the structure of the discussions. Similarly, the authors analyzed open bug reports to assess who contributes the reports, the frequency of resolution, and the patterns of comments between the bug reporters and the developers [21]. Though the results of these studies illuminate important elements of the design discussions in OSS, they say little about the consensus building process. Our work fills this gap by providing a comprehensive analysis of UI design discussions in OSS from the perspective of consensus building.

Thematic coherence and argumentation in OSS design discussions were analyzed in [4]. Their results showed the social influence of a user increases the depth of attention received in the ensuing discussion. Similarly, in [13], the authors studied how participation relates to code-related design decisions. One finding was that in the more effective projects studied, the number of participants increases over time and shifts from administrators to other community members. Our work builds upon these results by including variables related to participation and social influence, among many others, in our statistical analysis of consensus.

Studies of Group Decision Making and Consensus

There is a long history of research examining how various factors of a group such as size, task, and anonymity relate to decision making performance [3]. This line of inquiry has often compared technology-mediated and face-to-face (FTF) group work and relied primarily on the use of controlled studies. For example, controlled studies have shown it is more difficult and takes more time to reach consensus when groups use synchronous or asynchronous communication technology than when working FTF [17, 25]. Similarly, it is more difficult to reach consensus as group size increases [14] and, when consensus is required, group members experience less satisfaction when using communication technology than when working FTF [25].

While working FTF is ideal for reaching consensus, decision making in OSS communities often occurs in a distributed setting. It is therefore important to understand how consensus is achieved in this context. Our research fills this void by studying real world consensus decisions made in the UI design discussions in a mature OSS community.

Enabled by open access to peer production communities such as Wikipedia and OSS, researchers have begun to study similar elements of group decision making in real world data sets. For example, Lam et al. studied how group size, experience, and group formation influences decision quality in Wikipedia and found that larger groups make better decisions [22]. Similarly, Burke et al found that extensive and diverse contributions in Wikipedia can predict promotion decisions [7]. Analogous to these studies, our work tests how different factors relate to reaching consensus on UI design issues in a peer production community. Interviews were also used to understand users' perspectives on the consensus building process.

RESEARCH QUESTIONS AND METHODOLOGY

This research aimed to better understand the nature of consensus building in distributed discussions of UI design issues and centered on answering the following questions:

- How important is consensus building in these types of UI design discussions, what are the key challenges of reaching consensus, and what are the consequences of not reaching consensus from the user's perspective?
- What techniques are currently utilized for promoting consensus and how effective are these techniques?
- What factors affect consensus building? For example, how does the content (e.g. what is discussed), process

(e.g. number of design alternatives proposed), and social aspects (e.g. who participates) affect consensus?

These questions are not exhaustive, but are intended to offer initial understanding of consensus building in UI design discussions and identify opportunities for enhancing the discussion interface to promote consensus. To answer these questions, a mixed methods approach was employed consisting of semi-structured interviews and analysis of a large corpus of interaction data.

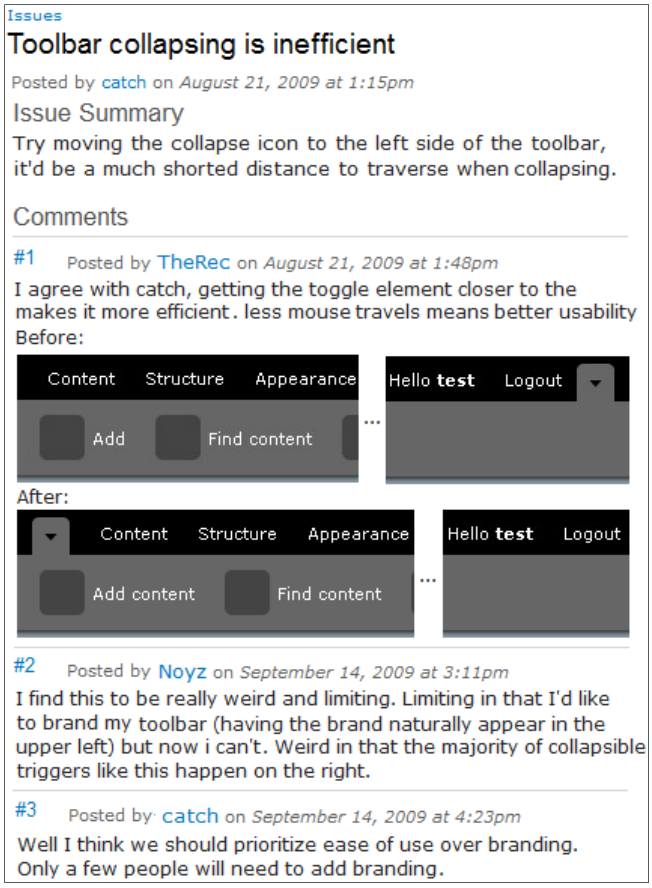

Figure 1. A UI design discussion occurring in the issue management system of Drupal. Participants are proposing and debating ideas for where to locate a shortcut on the toolbar.

User Interviews

We conducted 17 semi-structured interviews with designers and developers participating in either of two open source projects, Drupal and Ubuntu. Eight designers were interviewed, five from Drupal and three from Ubuntu, with an average of 4.5 years of experience in the community (σ=2.6). Nine developers were interviewed, six from Drupal and three from Ubuntu, with an average of 5 years of community experience (σ=2.6). Each interview lasted about an hour and was conducted via phone (n=14) or IM (n=3), whichever a participant preferred, and remuneration was either a $25 or $30 gift card depending on the duration of the interview. We will refer to Drupal and Ubuntu designers as DD# and DU# and Drupal and Ubuntu developers as DevD# and DevU#, respectively.

We first asked a participant to describe one or two recent or memorable discussions s/he participated in. In context of these discussions, we asked the participant to describe the consensus building process, what is hard about this process, the techniques utilized to foster consensus, the factors affecting consensus, and the consequences of not reaching consensus. Twelve interviews were conducted prior to the data analysis and five were performed afterward. For the latter interviews, a few questions were added to probe further about specific results of the analysis.

Interviews were coded to derive common themes using a Grounded Theory approach [26]. The results were used to gain insight into the consensus building process, identify features to include in our interaction analysis, and help interpret the results.

Collection of Authentic Interaction Data

The interaction data was extracted from the discussion threads (discussions) in the issue management system of Drupal, an open source content management system initiated in 2001. Drupal is a mature community with an established workflow and social organization. At the time of data collection, for example, the software was being used in about 490,000 websites to manage content and about 440,000 people had registered to contribute to the project.

Changes to the user interface and system software of Drupal are requested, discussed, and implemented (or not) through its issue management system. See Figure 1. Any community member can create an issue in the issue management system describing a design problem or feature request, which establishes a separate discussion. Others may participate in the discussion by proposing design alternatives, critiquing the alternatives, implementing an alternative (writing a patch), reviewing a patch, clarifying the problem, or offering other insights. To indicate the current progress of a discussion, participants can set its status to 'active', 'needs work', 'needs review', 'reviewed and tested by the community', 'fixed', or 'closed'.

There are four categories of discussions in the issue management system: bug reports, feature requests, tasks, and support requests. According to drupal.org, bug reports aim to resolve functionality and usability problems while feature requests are for adding new functionality. Tasks are non-functional things that 'need to be done' while support requests are for technical support. We only examined bug reports and feature requests as they contained the majority of the UI design discussions we wanted to study.

There were 285,008 discussions tagged as bug reports and feature requests in the issue management system at the time of data collection. This set was filtered to include only the discussions tagged with "Usability" or "d7ux" (usability in Drupal 7), which left 577 UI design discussions. These discussions occurred between March 2004 and September

2011. The usability issues ranged from significant redesigns to design details. For instance, an issue titled "*Initial D7UX admin overlay*" aimed to revamp the interaction design of admin pages in Drupal by providing themed admin pages as an overlay on top of the actual website while another issue only requested changing the location of a shortcut on a toolbar (see Figure 1).

We used the status to categorize discussions as consensus, non-consensus, or ongoing (unclear if consensus has been reached). We considered discussions marked as *closed* as having reached consensus. This typically means there was collective support for a decision such as implementing a specific proposal or concluding the issue was unnecessary or not a problem after all. This status is a reasonable proxy for consensus because if any participant strongly objected to the proposal, s/he could have reverted the status (e.g. back to 'active') and caused the discussion to continue.

Differentiating non-consensus and ongoing discussions was more challenging. We calculated the idle duration, the time from when the last comment was posted to the time of data collection and considered a discussion to be non-consensus if it's idle duration was more than 90% of the idle durations between comments in the consensus discussions. The remaining discussions were considered ongoing. This categorization yielded 284 consensus and 241 non-consensus discussions. The 52 ongoing discussions were discarded.

Finally, we filtered this data set to include only those discussions that were non-trivial. By reading a large fraction of the discussions and experimenting with different thresholds, we found that a threshold of having at least seven comments filtered almost all of the non-trivial discussions. After this filtering, we had 200 consensus threads and 141 non-consensus threads for our analysis. Table 1 reports summary statistics for the consensus and non-consensus UI design discussions after this filtering.

Characteristic	Type	Mean	SD
Discussion duration (weeks)	C	41.76	50.71
	NC	101.70	83.17
Number of comments	C	48.34	61.02
	NC	33.21	36.95
Number of participants	C	12.82	10.42
	NC	11.82	8.77

Table 1. Summary statistics for the consensus (N=200) and non-consensus (N=141) UI design discussions. 'C' indicates consensus and 'NC' indicates non-consensus discussions.

USER PERSPECTIVES ON CONSENSUS BUILDING

In this section, we first describe how consensus building in UI design discussions differs from technical discussions. Drawing from the interviews, we then report user perspectives on the challenges of consensus building, the benefits of consensus and the consequences of not reaching consensus, and techniques used for promoting consensus.

Consensus Building in UI design Discussions

Participants felt that consensus building in the UI design discussions is not the same as in the technical discussions (e.g. on performance and security) due to more subjectivity, lengthier discussions, greater visibility of the outcomes, and difficulty of presenting supporting evidence.

"*Usability also, is an area usually more people can get involved, some of the highly technical stuff not that many people can get involved in the issue without background. In usability it would be easier, anyone can suggest a design: I think we should do it like that. It tends to get a lot more people involved; it's also a lot more visible […]. Often more people are participating, more people means it's harder to [build] consensus.*" [DevD4]

For instance using discussions tagged with "Performance" as a sample of the technical discussions (N=322), we found that 703 unique participants contributed to the technical discussions while 1404 unique participants contributed to usability discussions (N=577). On average, more unique participants contributed to a usability discussion (μ=9.47, σ=9.45) than to a technical discussion (μ=7.98, σ=8.11; $t(897)=2.38$, $p<0.05$). We also found that participants spent more time in weeks discussing a usability issue (μ=57.99, σ=10.12) than a performance issue (μ=38.46, σ=49.33; $t(897)=4.42$, $p<0.05$). Having more people participate in a lengthier, subjective discussion is perceived to make consensus building more challenging.

Another factor that makes consensus building more challenging in UI design discussions is the difficulty of providing supporting evidence in the discussions. For instance, conducting usability testing to assess the effectiveness of a design proposal needs a lot of time and effort while testing the effectiveness of a performance solution can usually be performed directly and is therefore perceived to be easier:

"…I think the usability testing is a little harder to do often. It takes a little more time, if you do an informal one it's not so bad, but you certainly can't do with just one person, […] you have to get several, and it takes time to get evidence. In some other areas in Drupal it's easier to get evidence. You know, what percent [it makes the Drupal website] faster." [DevD4]

Benefits of Reaching Consensus and Consequences of Not Reaching Consensus

Participants stated that reaching consensus in the UI design discussions was critical for building a better product (n=4) and for strengthening the community (n=4). To both of these points respectively, one participant (DD4) explained:

"…when we reach consensus we are taking our strengths [to] make the world together, we have something that is at least as good as what the two of us could bring to

separately, and probably is better because our strengths tend to reinforce each other."

"…The more we can reach consensus in itself, the fact that we reach consensus in itself, foster stronger feeling in community, so the little instances of that build on one another and help us become stronger as a community, and therefore more likely to invest in reaching consensus on other projects, on other issues."

On the other hand, the inability to reach consensus can result in an unimproved product, build resentment in the community, and demotivate community members to the point of leaving the discussion or the community altogether.

"[Consequences of not reaching consensus are] stupid interfaces surviving yet another version in Drupal, known issues not being fixed, frustrated contributors. Consequences can be that people disappear for a couple of weeks or entirely because they get burned out on a too long discussion that didn't reach consensus… " [DD5]

The inability to reach consensus also causes the loss of significant community effort. For example, of the 577 UI design discussions we analyzed, 241 (42%) did not reach consensus. These discussions contained 4968 messages and 460 patches, contributed by 1934 participants. This outcome highlights the need for techniques for enhancing consensus building within the UI design discussions.

Challenges of Consensus Building

Despite recognizing the importance of consensus building participants identified key challenges that make consensus difficult to achieve. For example, one challenge is bridging the different perspectives and needs of the community members engaged in a UI design discussion (n=8):

"There are many different use cases for Drupal, what is optimal for one use case may be suboptimal for another, and there are strong differences of opinion within the Drupal community about which use cases, if any, should be given preference… Some people build for small sites, some people work on large sites, some people are designers, others are developers [and] others are end users" [DevD2]

Another challenge is overcoming a strong sense of ownership over one's contributions (n=4). For example, one reason that members contribute is because they can adapt the software to their own needs [15]. However, building consensus requires members to detach themselves from their own contributions and consider alternatives:

"People have egos and they have a lack of human contact with the people that they are talking to and trying to discuss with and a lot of time because these ideas are our own creations and our own feelings it's very difficult to separate ourselves from our own egos." [DevU1]

Similarly, expressing strong emotions during a discussion can also hinder consensus building (n=2). As DD4 said:

"certainly, there've been some that people just got so frustrated that their emotions, me included, to some extent let our emotions lead the way of communicating, rather than communicating based on the facts that matter…"

Other challenges for building consensus identified by interviewees included not having enough participants who are interested in a discussion (n=2), having too many participants lacking necessary background or general design knowledge (n=2), the absence of evidence supporting various claims (n=1), and lack of time and resources (n=1).

Techniques for Promoting Consensus

From the interviews, we identified different techniques that designers and developers use to promote consensus. Our interviewees said that providing evidence in support of a design proposal can better convince opposing parties in a discussion and accelerate consensus building (n=7). For instance, sharing the results of usability testing on a design proposal or showing how the proposal worked in a similar situation can convince other participants.

"what we do is look to what other projects and other web standards exist along the problem, say the problem of where to locate the help link, on the admin bar, we look at a bunch of different web applications, Facebook, Google docs, also desktop applications things like that, basically starts to conform a consensus around that ok it got to be on the right hand side of the menu and so we tend to go with that …" [DevD6]

Another technique for promoting consensus was presenting screenshots or writing a patch for a design proposal (n=5). As DevD2 said participants in a discussion are more likely to comment on a proposal that has a patch attachment:

"I present arguments in favor of it and then post a patch. People are typically more inclined to go with a solution that has a patch than another solution that does not have a patch, unless they have a major reason for liking the other solution better."

Participants also noted endorsing experienced members of the community in the discussion (n=3), writing a summary of the discussion (n=3), communicating via synchronized channels (n=3), having an administrator make the final decision (n=5), spending time to understand others' perspectives (n=3), voting for different design proposals (n=2), and advertising a stalled discussion (n=2) as techniques for promoting consensus. However, it is unclear how effective these techniques are given that 42% of the UI design discussions we analyzed did not reach consensus.

INTERACTION DATA ANALYSIS

To understand quantitatively which factors correlate with consensus, we analyzed the interaction data collected from the Drupal issue management system as described earlier. Based on the interview results, consensus building literature [6], and prior analyses of online communities [1, 2], we identified 23 metrics that may relate to consensus building.

Category	Metrics
Content	(1) Avg. # of words *
	(2) Total # of words
	(3) Priority of the issue
	(4) # of "usability testing"s
	(5) # of "summary"s *
	(6) # of screenshots *
	(7) # of "code review"s *
	(8) # of non-Drupal links
	(9) # of "IRC"s
	(10) # of "?"s *
Process	(11) Duration of the thread *
	(12) Avg. duration between comments
	(13) # of comments *
	(14) # of patches *
	(15) # of contributors *
	(16) # of comments by thread initiator *
User Relationships	(17) # of triads in the social graph
	(18) Avg. page rank score of contributors
	(19) Avg. # of total participation duration
	(20) Avg. # of contributors' prior comment *
	(21) # of participation weeks of initiator *
	(22) # of prev. comments of initiator *
	(23) # of alternate replies *

Table 2. The metrics used in our regression analysis. The metrics marked with (*) were later removed from the analysis to avoid problems with collinearity.

Table 2 lists these metrics grouped into three categories: content, process, and user relationships. Though not exhaustive, these metrics provide a useful starting point for understanding which factors affect consensus building in UI design discussions.

For example, for Content metrics, we counted the number of messages with screenshots attached as a proxy for the number of design alternatives proposed. More alternatives may create more opportunities for consensus. The number of question marks was counted as a proxy for attempts at building shared understanding [5]. From the interviews, we found that synchronous chats can promote consensus and therefore included how often "IRC" was mentioned in a discussion. Similarly, occurrences of "usability testing", "code review", and "summary" were counted in each discussion. The number of non-Drupal links was included to capture use of external evidence in the design arguments.

For the Process metrics, we counted patches, comments, and contributors as a proxy for the level of activity in each discussion. A discussion with increased activity may have a better chance of reaching consensus. The duration of a discussion was also included as allocating more time to a discussion may indicate stronger commitment to identifying an agreeable solution.

For User Relationships, we calculated the number of triads contributing to a discussion to estimate prior interaction history [12]. Triads were determined from the social graph created from the users, discussions, and relationships [11, 12, 17]. In a social graph, the nodes represent users and discussions while edges represent their relationships. An edge between a user and discussion is established when a user contributes to that discussion. An edge between two users is established when one user responds to the other. An edge is weighted based on the length of the comment. We also computed a page rank score [24] for each participant to estimate 'influence' within the community. The page rank score was also calculated from the social graph. The duration of community participation was used to estimate the experience of participants as interviewees felt having more experienced members of the community participate in a design discussion promotes consensus.

To calculate values for these metrics, we incorporated information from the discussion content (e.g. length of comments), metadata of the discussions (e.g. duration of the thread), and contributor's Drupal profile.

Logistic Regression

To investigate how these metrics relate to consensus, we performed a binary logistic regression. Binary logistic regression is a type of regression used to model the relationship between independent variables and a binary response variable. For our analysis, the metrics from Table 2 served as the independent variables and were computed for each discussion in our data set while the dependent variable was whether the discussion reached consensus. To avoid problems with collinearity in the regression analysis, we removed fourteen variables that demonstrated strong correlations ($r>0.4$). These variables are marked in Table 2.

We performed binary logistic regression as implemented in SPSS and used step-down regression to identify our partial model. We first entered all variables and removed each variable that did not show significance and repeated until a set of variables was reached that were all significant. Three of the nine metrics included in the analysis showed significance ($p<0.05$): average number of participation weeks, number of triads, and mentions of IRC. Table 3 summarizes the results. To assess the goodness of fit of our model, we performed the Hosmer-Lemoshow test (Chi-square=7.91.56, p=0.44). In this test, the model is valid if the p-value is *greater* than 0.05.

To aid interpretation of the results, we conducted five follow-up interviews as described in Methodology. These interviews followed the original script, but probed further about the factors found to be significant. In addition, we analyzed thirty of the discussions that reached consensus in

	B	Df	Sig.	Exp(B)
Average # of participation weeks	.01	1	.00	1.01
# of triads in social graph	.10	1	.03	1.10
IRC	.24	1	.05	1.26
Constant	-1.06	1	.02	.35

Table 3. Results of the binary logistic regression. The Hosemer-Lemoshow test confirmed the validity of our regression model (Chi-square=7.91, p=0.44).

our data set. The discussions were sorted based on the three factors found to be significant and ten threads from the top of each of these three lists were reviewed.

Experience with Drupal

Our regression analysis showed that having people in a discussion who have participated in Drupal longer promotes consensus. Research studies confirm that including experienced people can positively influence group decision making performance [19]. Our interview results and review of the discussion threads illustrated how experience can facilitate consensus building. First, we learned that members who have been in the community for a long time facilitate consensus by helping other members, especially new ones, understand the norms of communication and the process of participation in the community.

"…what's important for reaching consensus is having common ground rules or communication and process and you know working those out and to that extent more experienced in Drupal community might help people be better at reaching consensus, because they'll understand that you don't say things this way or you do, or these are the options for contacting somebody if you have a problem or that kind of thing…" [DD4]

Second, it was reported that experienced members are more skilled in unblocking a discussion. For example, comments and opinions posted by experienced members are valued more than those posted by other participants.

"*…there are those people in the community that are recognized – people who have been in the community a long time, or who are respected because they have written a lot of code, or they have written a lot of patches, or they are the maintainer of a certain bit of code – and when those people chime in, it tends to hold a little bit more weight when someone unknown chimes in*" [DD1]

As a result, experienced members can help direct the discussion toward a specific design proposal. For instance, in a discussion about adding edit and delete operations to a page in Drupal, when two of the participants (X and Y) proposed different solutions and were not able to come to an agreement, a community member with design experience was invited to review and decide between the proposals:

"I am inclined to agree with X here, following the logic of menus and taxonomies this should make more sense…"

Finally, experienced members can promote consensus by understanding the need for proposing solutions that accommodate competing alternatives. Satisfying opposing views allows stalled discussions to move forward. For instance, in a discussion about placement of a shortcut for collapsing the Drupal toolbar, X thinks that the icon for the shortcut should be placed on the left side of the menu to prevent accidental clicks on the "logout" icon while Y thinks it should remain on the right side because the space on the left is needed for branding. They cannot come to an agreement until Z who has been in the community for six years comes in and proposes a new solution:

"Thought: Move /help over to the right of "log out", move the shortcut collapsing back to the right, then you'd at least accidentally click a "safe" link."

Prior Interactions

Our analysis showed that having more triads participate in a discussion increases the likelihood of consensus. Triads represent three people who have previously interacted and produce closed social structures that promote trust [29].

Interview results and review of the discussions confirm trust as an important factor in consensus building. First, we found that participants are more likely to read, learn from, and evaluate comments posted by members whom they trust. This exchange of knowledge can create mutual understanding and consequently promotes consensus.

"*I think I'm less likely to dismiss something if it's from somebody I know and I respect. It's a little more likely to read carefully what they say and believe that they have something meaningful to say*" [DevD4]

This finding reflects findings in other research studies that indicate a high degree of trust existing within dense parts of a social network facilitates the exchange of complex knowledge [31]. Second, prior interaction and increased trust promotes agreement among participants.

"… it's sort of like a trust matrix type of thing, because if I don't know you and you are suggesting this thing that sounds like a bad idea to me, I probably fight against it, but if you are proposing something and I don't know you but three other people that I do know are saying yeah actually that's a great idea and this is why, then I'll be far more likely to be like alright let's go with it then."[DevD6]

Finally, we found that trust in other participants' technical abilities can save time in the process. For instance, knowing that the person who wrote a patch usually conforms to coding standards can accelerate code review.

"… if people know who somebody else is, it saves a heck of a lot of time, at all levels, like, for example, if I know the person who wrote the patch and I know that traditionally they write pretty good patches that conform to coding standard and stuff like that and then I see the person that reviewed it is the person I associate with being the smart person about that thing and the person who marked it as

reviewed and tested by community […], and that person was also someone I recognize as if they say something is RTBC it's actually good to go. Then it saves all kinds of time." [DevD6]

Similarly, recognizing the person who wrote the patch as a skillful programmer can accelerate implementation. For instance, in a discussion where the proposal was to add an edit link to all Drupal pages, one member (Y) was able to build upon another's (X's) patch and save time. Y says:

"So, yeeha, X's last changes contained some really good ones that allowed me to proceed further."

Use of Synchronized Communication Media

Based on the regression analysis, threads containing more mentions of "IRC" are more likely to reach consensus. A group of two to five people usually participate in the synchronized discussions and are expected to report their conclusions back to the corresponding discussion for the benefit of all. Failure to report may cause the other participants to loose context.

"*The danger in IRC becomes when and this happens sometimes when there are huge discussions that go on IRC, big community impacting discussions and only the people who happen to be on IRC at that time, know about them and if those don't make their way back to the issue queue or groups or some other mean of more permanent storage that's really dangerous because a lot of people loose context in these discussions that way*" [DevD6]

From the interviews and our inspection of discussions, we identified three ways that IRC can help build consensus. First, we found having discussions in IRC can accelerate agreement between opposing viewpoints.

"… it [IRC] can help if there is one or two people who are disagreeing about something, if those people go to IRC they can chat it out much faster than the issue queue." [DD4]

Second, in IRC people can come up with an initial design proposal for solving the usability problem that may not be possible in context of the larger discussion.

"*IRC is great for say a small group of people going off and coming up with an initial proposal that they all agree on and then proposing that to the community.*" [DevD6]

Reporting this proposal back to the discussion advances consensus building because participants can argue for or against the proposal as opposed to developing their own. As DevD6 said:

"…then it becomes let's argue against this position as opposed to try to come to a position to argue against…"

Finally, we learned that participants use synchronized communication to hasten collaborative design review, programming, and debugging sessions. This finding corroborates observations reported in [10]. For example, when discussing the design and implementation of an overlay for the Drupal interface, one of the developers (Y) asked another developer (X) to join him in a synchronized chat for a collaborative debugging session:

"*X, some of your files are being cut off, such as overlay-parent.css(?) Please come onto IRC so we can help you debug.*"

DISCUSSION

Our regression analysis showed that three of the factors tested are predictive of consensus in a UI design discussion: the experience of participants, number of triads, and mentions of synchronous communication. Interestingly, none of the content metrics were significant. One interpretation of this result is that who participates in a UI design discussion is more important than how many design alternatives are proposed or what arguments are made for the purpose of building consensus. For example, this may be due to not having a facilitator in the discussion skilled at steering the group toward consensus [8]. Participation of experienced members may therefore compensate for the absence of trained facilitators, i.e., they have a better understanding of how to guide the discussion toward consensus. Another possibility is that the content metrics used in our analysis were incomplete. Future work should therefore examine additional metrics such as the use of different argument types [23] and rhetorical devices [20] to further test how content attributes may relate to consensus.

A number of factors perceived by our interviewees to relate to consensus did not show significance in the regression analysis. For instance, interviewees mentioned contributing concrete evidence to an ongoing discussion such as usability tests of the design proposals and external links to interface examples positively affect consensus building.

One reason these factors did not correlate with consensus is that they were seldom performed. For example, in our data set, mentions of usability appeared in only 0.06% of the consensus discussions and in 9% of the non-consensus discussions. One way to foster the inclusion of concrete evidence is to establish specific community guidelines for discussing UI design issues. Another method would be to configure a testing platform where participants can easily try a patch and provide feedback in the discussion without having to worry about applying the patch to their locally installed version of the product.

A second possible reason some of the factors did not show significance in the analysis is that we did not consider their context. For example, in our data set, the number of links to external sources was similar in consensus (μ=5.5, σ=7.5) and non-consensus (μ=4.9, σ=6.9) discussions. This may be due to not considering the helpfulness of the link targets. For example, future work may consider weighing the link counts based on the helpfulness of the link targets to the discussion (e.g. did the links reference sketches of design proposals, interaction examples within well-known Web sites, or community design standards).

Design Implications

Our work has several design implications for discussion interfaces w.r.t. promoting consensus. One implication is to enable discussion participants to quickly identify others with whom they have had prior interactions. These community members could then be invited to join the discussion, thereby increasing the number of triads. For example, for each discussion, the community software could maintain a list of members whose participation would form triads by analyzing the social graph [18] or history of participants' contributions [9]. Options could be offered for filtering the list, e.g., requiring a minimum number of prior interactions or specifying that only the interactions within specific types of design discussions be considered.

A related implication is to allow discussion participants to identify experienced members who may be willing to join the discussion. Inviting appropriate people to join a discussion may not only aid consensus building, but may also assist community members in identifying discussions of interest. For example, analogous to [9], the system could recommend experienced members appropriate for the discussion by considering the duration of their community membership, interest profiles, and recent activity within the community (to prevent core members from receiving too many invitations). As before, options could be provided for modifying these search parameters.

Our analysis showed that participants value the comments contributed by experienced members or members with whom they have had prior interaction. The discussion interface could therefore allow participants to filter comments within the current discussion contributed by others meeting these criteria or by including appropriate visual cues for these criteria within the comments.

Results of our interviews and inspection of discussions revealed that certain types of comments aid consensus building more than others. For example, comments that strongly argue for or against design alternatives can build agreement, comments that summarize the discussion to date can help participants make sense of the thread, and comments that report the conclusions from synchronized discussions can help participants maintain context. The discussion interface could therefore employ color codes or other visual cues to highlight these types of comments [24, 30]. To classify comments, the author or other participants could be allowed to assign pre-defined community tags. To reduce or eliminate the costs of tagging, an alternative would be to automatically infer the comment types, which could be modified by participants to correct any errors.

To further aid the consensus building process, recent key contributions to the discussion could also be highlighted. For instance, comments that include key contributions such as the most recent design proposal, implementing or reviewing a recent patch, or changing the status of the discussion could be highlighted. It is important to note that not all of the filtering, searching, and highlighting mechanisms described need to be included in a discussion at the same time. For example, end users could configure which of the features are applied in their local interface.

Limitations

One limitation of this work is that our regression analysis did not include an exhaustive set of factors that may relate to consensus. Additional metrics such as the language complexity of the messages, number of arguments for or against design proposals, the sentiment of those arguments, use of rhetorical devices, and advanced techniques for assessing expertise could be included in future analyses.

Second, our qualitative findings were derived from the responses of seventeen participants. We look forward to collecting data from additional community members and analyzing how the responses relate to their different roles such as designer, developer, administrator, or end user.

Finally, our interaction data was collected from the UI design discussions in one open source software community. Similar analyses should be performed for UI design discussions in other open source and distributed software projects to assess the generalizability of our results.

CONCLUSION

Consensus building is a critical component of UI design discussions in OSS as it promotes a better product and a stronger community. In this paper, we studied consensus building in UI design discussions from an established OSS community using qualitative and quantitative methods. Our results made three contributions. One contribution was reporting user perspectives on the challenges of reaching consensus in UI design discussions, the techniques utilized for addressing the challenges, and the consequences of not reaching consensus. A second contribution was analyzing how various metrics related to the content, process, and user relationships of the discussions correlate with reaching consensus. The main result from this analysis shows that discussions having participants with more experience and prior interaction history are more likely to reach consensus. Finally, we offered design implications for promoting consensus in distributed discussions of UI design issues.

One immediate direction for future work is to implement our design implications and test their utility and impact on the consensus building process. A second direction is to analyze the logs of synchronized communications that occurred within the discussions studied to better understand the strategies used by participants and the effect of this medium on consensus. Finally, one can conduct a similar study on technical discussions (e.g. on performance and security) to shed more light on the differences between UI design and technical discussions for consensus building.

ACKNOWLEDGEMENT

This work was supported in part by the National Science Foundation under award no. IIS 06-43512.

REFERENCES

1. Adamic, L. A., Zhang, J., Bakshy, E. and Ackerman, M. S. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *Proc. WWW*, 2008, 665-674.
2. Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G. Finding High-Quality Content in Social Media. In *Proc. WSDM*, 2008, 183-194.
3. Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C. and LaGanke, J. S. Computer-Mediated Communication and Group Decision Making: A Meta-Analysis. *Organizational Behavior and Human Decision Processes*, 87, 1 (2002), 156-179.
4. Barcellini, F., Détienne, F., Burkhardt, J.-M. and Sack, W. A Study of Online Discussions in an Open Source Software Community. *Communities and Technologies*, 2005, 301-320.
5. Benkler, Y. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale Univ. Press, 2006.
6. Briggs, R. O., Kolfschoten, G. L. and Vreede, G.-J. d. Toward a Theoretical Model of Consensus Building. In *Proc. Americas Conference on Information Systems*, 2005.
7. Burke, M. and Kraut, R. Mopping Up: Modeling Wikipedia Promotion Decisions. In *Proc. CSCW*, 2008, 27-36.
8. Clawson, V. K., Bostrom, R. P. and Anson, R. The Role of the Facilitator in Computer-Supported Meetings. *Small Group Research*, 24 (1993), 547-565.
9. Cosley, D., Frankowski, D., Terveen, L. and Riedl, J. SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia. In *Proc. IUI*, 2007, 32-41.
10. Erickson, T., Kellogg, W. A., Laff, M., Sussman, J., Wolf, T. V., Halverson, C. A. and Edwards, D. A Persistent Chat Space for Work Groups: The Design, Evaluation and Deployment of Loops. In *Proc. DIS*, 2006, 331-340.
11. Farnham, S., Chesley, H. R., McGhee, D. E., Kawal, R. and Landau, J. Structured Online Interactions: Improving the Decision Making of Small Discussion Groups. In *Proc. CSCW*, 2000, 299-308.
12. Fiol, C. M. Consensus, Diversity, and Learning in Organizations. *Organizational Science*, 5, 3 (1994), 403-420.
13. Hackman, J. R. and Katz., N. Group Behavior and Performance. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), Handbook of Social Psychology (5th ed.). New York: Wiley, 1208-1251.
14. Hare, A. P. A Study of Interaction and Consensus in Different Sized Groups. *American Sociological Review*, 17, 3 (1952), 261-267.
15. Hars, A. and Ou, S. Working for Free? Motivations for Participating in Open-Source Projects. *International Journal of Electronic Commerce*, 6, 3 (2002), 25-39.
16. Heckman, R., Crowston, K., Eseryel, Y. U., Howison, J., Allen, E. and Li, Q. Emergent Decision-making Practices in Free/Libre Open Source Software (FLOSS) Development Teams. In *Proc. IFIP International Conference on Open Source Software*, 2007.
17. Hiltz, S. R., Johnson, K. and Turoff, M. Experiments in Group Decision Making Communication Process and Outcome in Face-to-Face Versus Computerized Conferences. *Human Communication Research*, 13 (1986), 225-252.
18. Horowitz, D. and Kamvar, S. D. The Anatomy of a Large-scale Social Search Engine. In *Proc. WWW*, 2010, 431-440.
19. Kerr, N. L. and Tindale, R. S. Group Performance and Decision Making. *Annual Review of Psycholo*gy, 55 (2004), 623-655.
20. Ko, A. J. and Chilana, P. K. Design, Discussion, and Dissent in Open Bug Reports. In *Proc. iConference,* 2011, 106-113.
21. Ko, A. J. and Chilana, P. K. How Power Users Help and Hinder Open Bug Reporting. In *Proc. CHI*, 2010, 1665-1674.
22. Lam, S. K., Karim, J. and Riedl, J. The Effects of Group Composition on Decision Quality in a Social Production Community. In *Proc. GROUP*, 2010, 55-64.
23. Lemus, D. R., Seibold, D. R., Flanagin, A. J. and Metzger, M. J. Argument and Decision Making in Computer-Mediated Groups. *Journal of Communication*, 54, 2 (2004), 302–320.
24. Luther, K., Counts, S., Stecher, K. B., Hoff, A. and Johns, P. Pathfinder: An Online Collaboration Environment for Citizen Scientists. In *Proc. CHI*, 2009, 239-248.
25. Straus, S. G. and McGrath, J. Does the Medium Matter? The Interaction of Task Type and Technology on Group Performance and Member Reactions. *Journal of Applied Psychology*, 79 (1994), 87-97.
26. Strauss, A. L. *Qualitative Analysis for Social Scientists*. Cambridge University Press, 1987.
27. Susskind, L., McKearnan, S. and Thomas-Larmer, J. *The Consensus Building Handbook*, Thousand Oaks, CA: Sage Publications, 1999.
28. Twidale, M. B. and Nichols, D. M. Exploring Usability Discussions in Open Source Development. In *Proc. HICSS*, 2005.
29. Uzzi, B. Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. *Administrative Science Quarterly*, 42, 1 (1997), 35-67.
30. Willett, W., Heer, J., Hellerstein, J. and Agrawala, M. CommentSpace: Structured Support for Collaborative Visual Analysis. In *Proc. CHI*, 2011, 3131-3140.
31. Zaheer, A. and Bell, G. G. Benefiting From Network Position: Firm Capabilities, Structural Holes, and Performance. *Strategic Management Journal*, 26, 9 (2005), 809-825.
32. Zilouchian Moghaddam, R., Bailey, B. and Poon, C. IdeaTracker: An Interactive Visualization Supporting Collaboration and Consensus Building in Online Interface Design Discussions. In *Proc. INTERACT*, 2011, 259-276.