Shin, Joongi; Hedderich, Michael; Lucero, Andres; Oulasvirta, Antti

Chatbots Facilitating Consensus-Building in Asynchronous Co-Design

Published in:
UIST '22: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology

DOI:
10.1145/3526113.3545671

Published: 28/10/2022

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Shin, J., Hedderich, M., Lucero, A., & Oulasvirta, A. (2022). Chatbots Facilitating Consensus-Building in Asynchronous Co-Design. In M. Agrawala, J. O. Wobbrock, E. Adar, & V. Setlur (Eds.), *UIST '22: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (pp. 1-13). Article 78 ACM. https://doi.org/10.1145/3526113.3545671

Chatbots Facilitating Consensus-Building in Asynchronous Co-Design

Joongi Shin
Aalto University
Espoo, Finland
joongi.shin@aalto.fi

Michael A. Hedderich
Aalto University
Espoo, Finland
michael.hedderich@aalto.fi

Andrés Lucero
Aalto University
Espoo, Finland
lucero@acm.org

Antti Oulasvirta
Aalto University
Espoo, Finland
antti.oulasvirta@aalto.fi

ABSTRACT

Consensus-building is an essential process for the success of co-design projects. To build consensus, stakeholders need to discuss conflicting needs and viewpoints, converge their ideas toward shared interests, and grow their willingness to commit to group decisions. However, managing group discussions is challenging in large co-design projects with multiple stakeholders. In this paper, we investigate the interaction design of a chatbot that can mediate consensus-building conversationally. By interacting with individual stakeholders, the chatbot collects ideas to satisfy conflicting needs and engages stakeholders to consider others' viewpoints, without having stakeholders directly interact with each other. Results from an empirical study in an educational setting (N = 12) suggest that the approach can increase stakeholders' commitment to group decisions and maintain the effect even on the group decisions that conflict with personal interests. We conclude that chatbots can facilitate consensus-building in small-to-medium-sized projects, but more work is needed to scale up to larger projects.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

Co-Design, Chatbot, Consensus-building

ACM Reference Format:
Joongi Shin, Michael A. Hedderich, Andrés Lucero, and Antti Oulasvirta. 2022. Chatbots Facilitating Consensus-Building in Asynchronous Co-Design. In *The 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22), October 29-November 2, 2022, Bend, OR, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3526113.3545671

1 INTRODUCTION

Co-design is a class of design methods in which the diverse stakeholders in the design process (e.g., end-users) are actively involved as participants throughout the project, from the early stage of identifying needs to design and evaluation [51]. It includes collaborative ideation and design activities not only between designers but also between non-designers with shared accountability for the decisions. In short, in co-design, design is a *collective process.* The approach is argued to improve the end results' quality while also making design more democratic. This paper contributes to research on interactive systems for facilitating *co-design* [12, 44, 47, 60].

We look at interactive means for resolving conflicts and achieving consensus in a co-design project. *Conflict* is intrinsic to co-design. Conflicts naturally emerge between stakeholders' interests and in their understanding of each other because stakeholders bring varied expertise and experience. Without consensus, conflicts can remain, either implicit or acknowledged, and the stakeholders themselves may thwart a successful project outcome [1, 48, 52]. *Consensus-building* refers to attempts at converging views toward a collectively agreed decision [2, 5, 53]. However, achieving consensus does not mean that everyone's needs are addressed, which is often impossible. Rather, consensus involves agreement among stakeholders that the current group decision is a satisfactory one given the constraints, such as time and other resource limits [6]. Therefore, central to consensus is the stakeholders' willingness to commit to a group decision even if that decision is not fully satisfying. This is assumed to be achieved interactively in a process where stakeholders express personal needs and viewpoints, identify conflicts, discuss matters, and converge towards common decisions [7, 33].

Building consensus is challenging, though. It necessitates organizing a forum or an occasion that invites all stakeholders to meet and exchange views, regardless of differences in hierarchy or group size. It requires creating a sense of group effort as well [10]. In the context of co-design, such high effort and responsibility are expected of *moderators* who facilitate group discussion and co-design activities [23, 50]. Therefore, a risk of failing to build consensus exists if inexperienced moderators are in charge and cannot devote enough attention to individual stakeholders.

In this paper, we ask whether *conversational agents* could facilitate consensus-building in co-design projects. Conversational agents can indeed be instrumental to improve continuous interaction among stakeholders, as human moderators cannot. Our working hypothesis is that chatbots can work as asynchronous moderators that collect stakeholders' needs, identify conflicting needs,

UIST '22, October 29-November 2, 2022, Bend, OR, USA

ACM ISBN 978-1-4503-9320-1/22/10.
https://doi.org/10.1145/3526113.3545671

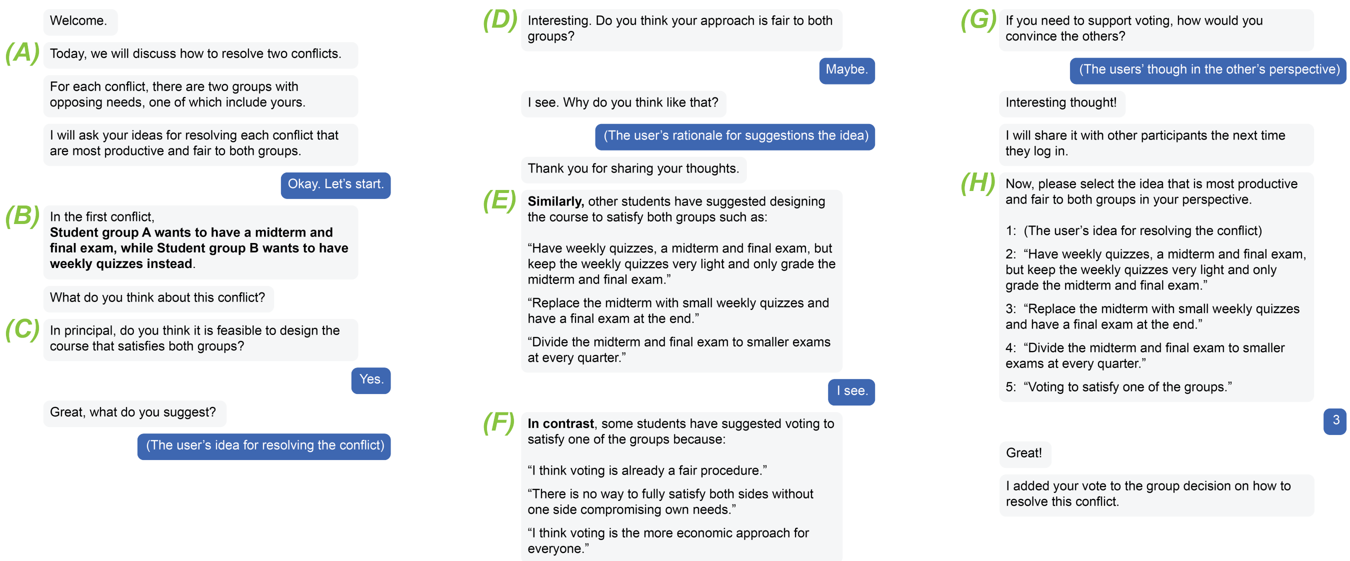

Figure 1: An example of chatbot-facilitated consensus-building in co-design that involves no direct communication between stakeholders. In the system, the chatbot stages the discussion (A) and presents conflicts (B). It then invites users to make an initial suggestion (C), perform self-assessment (D), review others' suggestions that are similar to (E) and in conflict with the user's own (F), take others' perspective (G), and make the final suggestion (H).

and try to resolve them via dialogue. Chatbots can be the bridge between stakeholders in this process, not only sharing each person's opinions but also managing the co-design activities and enhancing the trust among stakeholders required for building consensus.

A benefit of this approach is that it can make co-design *asynchronous*; that is, stakeholders can contribute their thoughts in their own time and space. The concept of asynchronous co-design has been proposed for bringing in a stakeholder who cannot attend co-design events in person [8]. The participants should be able to discuss their ideas without live communication among them. In such settings, we expect chatbots to hide individuals' identity to encourage them to share their true thoughts without worrying about criticism or about provoking uncomfortable social situations. Chatbots would also invite people who could have not attend co-design events physically and enable ideation at individuals' own pace. Accordingly, the scale of co-design projects can grow with improved depth and variety of information.

The challenge we focus on is *how to boost stakeholders' willingness to accept group decisions without them being in direct communication with each other*. We built our system based on the guidelines for building consensus through conversation, which have not been applied to co-design before this paper [6, 10]. In addition to facilitating idea exchange among stakeholders, all members should be able to sense group effort toward shared interest and believe their suggestions will not be ignored by those with other opinions. In particular, we proceeded from the principle that stakeholders should be motivated to follow group decisions without coercion. The process of conflict resolution needs to be considered fair and productive for all members, hence raising their willingness to commit to group decisions. Accordingly, we designed the conversation flow of a chatbot that leads the interaction with individuals. The chatbot asks what will benefit the group, presents other members' ideas, requests seeing ideas from others' perspective, and asks individuals to contribute to group decisions by selecting one of the ideas. Although prior work has looked at how conversational agents in online discussions can facilitate problem-solving and reduce effort [28, 32], how to facilitate asynchronous consensus-building in large collaborative efforts such as co-design has remained an open question.

We evaluated the approach in a study where we asked 12 university students to interact with our chatbot to resolve conflicts around the design of course activities. By comparing their responses before and after the interaction, while also considering their reactions to group decisions aligned with and conflicting with their interests, we addressed two research questions:

- RQ 1: Can rule-based chatbots increase users' willingness to commit to group decisions without direct interaction among the members of the group?
- RQ 2: Can rule-based chatbots increase users' willingness to commit to group decisions that conflict with personal opinions?

RQ2 articulates a particularly important process of co-design, in that consensus-building is often followed by other co-design activities (e.g., refining group opinions). The success of the chatbot interaction hence must be assessed in terms of stakeholders' willingness to commit even to conflicting group decisions, which can expect their continuous contribution in the later phases of co-design.

Our results indicate that interacting with a chatbot can indeed increase stakeholders' willingness to commit to group decisions by eliciting the perceived joint effort and fairness, also in the face of group decisions that conflict with one's own opinions. We discuss design implications for facilitating consensus-building with chatbots and the potential benefits that chatbots may bring to asynchronous co-design.

2 RELATED WORK

2.1 Interactive Support for Consensus-Building

Scholars have explored interactive systems that elicit individuals' preferences and support exchanging opinions throughout the process of reaching consensus. To support consensus-building in long-term design projects, Moghaddam et al. developed an online design-discussion platform that tracks design alternatives and designers' arguments [64]. Grünbacher and Briggs investigated the design of a shared platform for voting on key conditions and prioritizing the winning ones in aims of developing a space for negotiation [16]. To support disagreements' clarification, Liu et al. developed a user interface for ranking alternative solutions via several criteria established within the group [35]. Their systems enable users to compare their opinions with the group's current ones, hence achieving opinion alignment in the group. Similarly, Unehara et al. proposed a design-support system that automatically generates new design candidates that represent compromise among individuals' votes on the initial set of designs [56]. The systems hold promise for supporting space- and time-independent consensus-building, with users who come to the platforms later being able to examine the progress and contribute on the basis of the group's current decisions. However, they all rely on users' active participation and do not touch the discussion itself. How, then, can systems be designed to actively moderate and guide users in asynchronous consensus-building?

2.2 Interactive Support for Co-Design

While stakeholders contribute to co-design projects as experts in their respective domains, they often encounter difficulties to fully articulate or communicate their ideas comprehensively. Diverse co-design methods, techniques, and events have arisen in response to this issue [37]. Some employ tangible objects, such as physical probes [3] and cards [18]). Similarly, interactive systems have been designed to enhance stakeholders' involvement and ideation during co-design. For instance, studies have looked into affinity diagramming in a digital space, which allows swift rearrangement of thoughts to present multiple perspectives in converging ideas, not limited by physical boundaries [31, 36]. Mazalek et al. demonstrated the use of a digital tabletop to support storytelling among stakeholders in a local area [41]. With a focus on gaming environments, Walsh et al. facilitated co-design between children and designers by attending to children's familiarity with interactions in the digital world [59]. Enhancing concurrent co-design workshops has been a major focus on interactive systems. Little attention has been given to facilitating asynchronous co-design.

In response to the COVID-19 pandemic, Kennedy et al. explored translating multiple phases of physical co-design to a digital environment through online surveys and Zoom meetings [26]. Their study highlighted the benefits of resolving the power imbalance between researchers and other participants, alongside the setting's convenient control of discussion groups' size, whereby participants obtain more comfortable and intimate environments for sharing their thoughts. Whereas such work has proven to overcome physical limitations affecting co-design, there has not yet been enough attention to time-related limitations, such as reducing information disparities between people who join the co-design effort earlier vs. later. By adding chatbots to the process, we investigated the potential of facilitating asynchronous contributions in co-design at a convenient time for participants.

2.3 Chatbots as Moderators

To enhance stakeholders' communication and knowledge-sharing, various methods have been proposed, among them storytelling [23], brainstorming [63], contextmapping [57], future workshops [27], and embodied narratives [13]. All require the presence of moderators and hence bring in factors directly related to the limits of human moderators' abilities, such as the number of stakeholders or the frequency and duration of the co-design activities that moderators can manage. If these burdens can be lifted via interactive tools or artificial agents, a new design space for co-design projects could emerge.

Conversation is one of the most familiar channels in human communication during idea-sharing. Studies attest that more valuable information gets produced and that users engage more actively when the information is collected through conversational agents as compared to surveys [30, 49, 62]. In response, researchers have investigated these agents' potential as a replacement for human moderators to support design or discussion [17, 19, 46, 60]. For instance, to assist users in organizing their thoughts before joining group discussions [14] or engaging in persuasive writing [61], private discussion with a chatbot may be promising. By helping users hone the presentation of their personal ideas and rationales, chatbots have increased users' awareness of critical-thinking and argumentation skills. The studies closest to our area of interest were done by two research groups that investigated chatbot moderation in online discussions: In the study by Kim et al., a chatbot managed discussion time, encouraged "lurkers" to speak up, and organized users' opinions in real-time synchronous group discussion [28]. The authors also investigated chatbots for consensus-building, but the setting remained involving real-time discussion in small groups [29]. The other group, Lee et al., investigated chatbots' moderation of asynchronous online discussion [32]. The chatbot prompted users to input their thoughts on a shared platform and led discussions following a pre-designed structure. In these studies, users were expected to reach consensus based on the group decisions' quality rather than their perceptions from interacting with the chatbot and the other users. Our study, in contrast, examined the effects of chatbot interaction on users' willingness to commit to group decisions irrespective of how consistent these are with their own interests, and we assessed the potential of a chatbot as a moderator for stimulating group effort in asynchronous co-design.

3 CONVERSATION FLOW DESIGN

Our chatbot's task in co-design is to a) resolve conflicts among stakeholders' needs and b) increase stakeholders' willingness to commit to group decisions without the need for direct communication among stakeholders (i.e., the chatbot interacts with individuals). The overall conversation flow is shown in Figure 1. Firstly, the

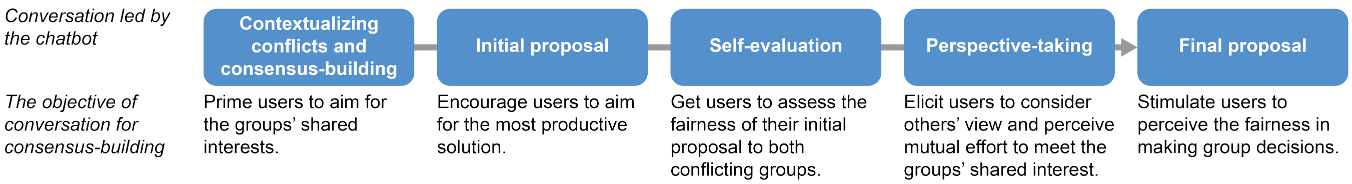

Figure 2: A conceptual diagram of consensus-building with our chatbot. By adopting the guidelines for building consensus, the chatbot elicits specific user behaviors that lead to greater user willingness to commit to group decisions.

chatbot introduces the purpose of the interaction and appropriate behaviors in building consensus. Then, it introduces a conflict and asks the user to suggest ideas for resolving that conflict. In the process that follows, the chatbot encourages the user to deeply think about the suggestions, by requesting self-evaluation and presenting potential solutions previously suggested by other group members. In the last part of the interaction, the chatbot displays all solutions from the current user as well as from the other participants in the co-design project. It then requests the user to select one solution that the user most prefers. After the user's input, the chatbot moves on to the next conflict.

The design for the conversation flow was based on guidelines for building consensus [10, 54]. Per these guidelines, the overall process of consensus-building follows the steps of defining the issue (conflicts), developing the criteria for consensus, proposing solutions, testing consensus, and making a final agreement. The entire process takes advantage of four basic beliefs among stakeholders that facilitate consensus-building. The first is that decisions should be in the group's interest. Everyone in the group should aim for the most productive and fairest solutions in resolving conflicts, targeting win-win conditions. The second belief is in devoting mutual effort to tackling conflicts as a group. Nurturing it evokes a sense of belonging, being in a group expanding effort for the shared interests. Hence, it stimulates active contributions. The third belief is in considering disagreements among stakeholders as offering chances to express different points of view. Finally, the belief that 'every voice matters' emphasizes fair opportunities for making suggestions and giving thoughtful consideration to each person's opinions, regardless of conflicting viewpoints.

During the interaction, our chatbot gets users to perform and perceive others' effort in accordance with the aforementioned beliefs fundamental to consensus-building. Figure 2 presents a conceptual diagram of consensus-building through our chatbot. Aiming for a win-win situation and caring about opposing viewpoints are addressed by directing users' attention accordingly. The chatbot challenges users to think about the most productive solutions (e.g., satisfying both sides in the conflict) before moving on to lesser solutions (e.g., satisfying one side only). A study by Kelly et al. applied a similar approach in the ideation phase of co-design to encourage 'daring' or 'impossible' solutions [25]. To prevent users from ignoring alternative opinions, the chatbot asks the user to try advocating the opposite ideas, in perspective-taking [22]. Compared to directly asking users for a certain input, having them perceive the presence of the others throughout the asynchronous interaction is a design challenge. Studying this subject, Narain et al. pointed to eliciting the sense of talking to other users by projecting the others' presence via chatbots [45]. Building on their work, we expected seeing the others' responses to stimulate a sense of mutual effort. The chatbot is programmed to show responses similar to the user's first, to enhance the sense of belonging, then opposing views, to reveal the possibility of resolving the conflict with different opinions.

4 EMPIRICAL EVALUATION

To observe the influence of the chatbot, we staged an asynchronous co-design project where teachers and university students design their course activities together. Specifically, we prepared a co-design scenario wherein stakeholders reach consensus on how to resolve conflicts between their opposite needs (e.g., student group A wants to have team projects while student group B does not want to have them).

4.1 Study design

We designed a within-group study to address our research questions. For **RQ1**, we followed the pre-post study procedure, comparing user responses to the same set of survey items right before and after interacting with the chatbot. In this study, it was critical to collect participants' baseline responses just before the chatbot interaction since their exposure to co-design might encourage them to intentionally display favorable behaviors for reaching consensus. By capturing participants' perceptions of their own willingness and their expectations of other members immediately before the interaction, we pinpointed the influence of our system, isolated from participants' latest understanding of co-design. For **RQ2**, we compared user responses from after seeing group decisions consistent with their personal interests (aligned group decision) and decisions counter to their personal interests (conflicting group decision).

The overall study structure is shown in figure 3. For examining the influence of the chatbot interaction, it was crucial that a) participants believe this to be a co-design project and b) participants be similar in their level of caring about conflicts. To provide the sense of an ongoing co-design process, we conducted the experiment over three days, simulating the time needed for collecting and analyzing all stakeholders' suggestions. The chatbot collected stakeholders' needs (day 1), discussed how to resolve conflicts (day 2), and revealed group decisions (day 3).

With an experiment of this duration, it is challenging to find mutual conflicts that are relevant to all participants. Therefore, we prepared personalized conflicts for each participant, based on the needs expressed by that participant on day 1. For each conflict, one of the needs came directly from that participant. We prepared the opposite need, along with three potential solutions for resolving

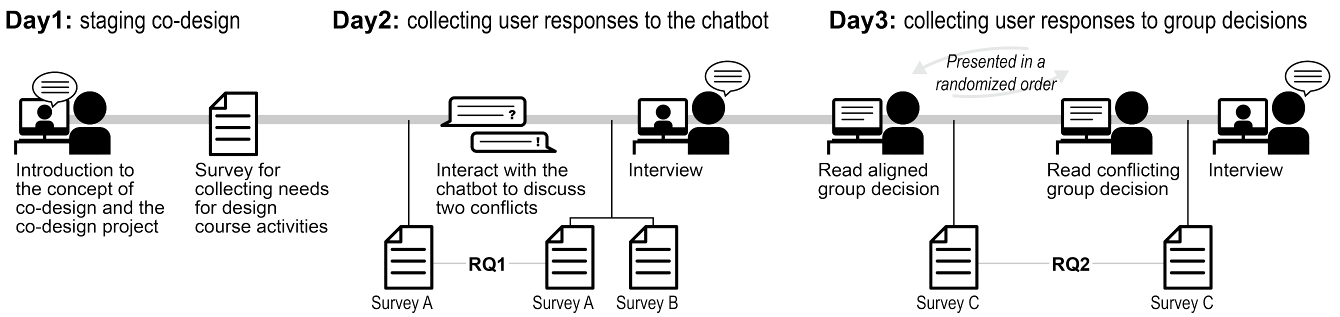

Figure 3: The structure of our empirical evaluation and the relationship between the research questions (RQ1 and RQ2). Survey A collected data on users' willingness to commit to group decisions, and survey B addressed the utility of the chatbot. Survey C collected user responses on the combined effect (A + B).

the conflict, which the chatbot presented as the others' suggestions (e.g. "The other students said…"). Likewise, aligned and conflicting group decisions were tailored to match each participant's input during the chatbot interaction.

4.2 Participants

We recruited 12 university students (mean age = 26.58, SD = 3.73; gender-balanced sample), from diverse departments, such as physics, chemistry, electrical engineering, computer science, and design. All participants had experience of taking part in team projects that involve resolving conflicts and building consensus through face-to-face group discussion, negotiations, making compromises, and voting. Half of the students had acted as a teaching assistant; nevertheless, we instructed everyone to participate as students. Four participants had some understanding of co-design prior to our study. To prime participants such that they shared similar levels of knowledge of co-design, we introduced the concept to all of them in line with a set script. Each participant was compensated with a restaurant voucher (25 EUR).

4.3 Task

Participants performed specific tasks on each day. On day 1, they filled in a survey form to share the needs that they wanted to consider in the design of course activities. On day 2, participants interacted with the chatbot without any interference from the research moderator. The participants took part in two discussions in a row, focusing on one conflict at a time. At the end of each discussion, the subject selected the solution that they want to try the most compared to the other potential solutions. Before and after the interaction, participants completed a survey, as noted above. On the last day, they were presented with aligned and conflicting group decisions. They shared their responses to each group decision via a survey and a post-interview.

4.4 Apparatus

We developed a rule-based chatbot to lead the conversation and nurture the aforementioned beliefs behind consensus-building, following the predefined order. Since we focused on what users perceived from conversation designed in advance, adaptive conversation that interacts with users' ideas further (i.e., NLP capability) was not required for this study.

Our conceptual conversation flow is independent of specific chat systems and chatbot software. We used the Telegram messaging platform (Figure 4). Thanks to its popularity [1], we could expect participants to be familiar with its interface or that of similar messaging services. Using an existing messaging platform also removed the need to set up a discrete communication system for the study. Our chatbot is written in Python and connects to the Telegram platform via an implementation of its API[2]. The participants connected to Telegram via a browser or desktop application, using their own accounts. Our chatbot's code is available online [3].

We implemented answer buttons for questions that required choosing from a closed set of options (e.g., 'Yes', 'Maybe', and 'No') and also items for which the user needed to select from among previously shown examples. The answer buttons freed the participants from typing repetitive answers and eliminated the risk of the chatbot misinterpreting a participant's answers to any questions. When the chatbot collected the participant's ideas and rationale, free-form input was used: the users typed in their answers to these questions.

4.5 Measurements

We argue that the quality of group decisions alone cannot clearly reveal the success of facilitating consensus-building. In principle, participants may go along with even a low-quality group decision if they believe it to be the best one at the moment [10]. Therefore, we probed participants' perceptions of the overall chatbot interaction.

We collected participants' responses through self-reporting surveys and post-interviews. The survey before and after the interaction used the same questionnaire, collecting participants' views about their willingness and that of the other participants in the co-design project. To measure the utility of the chatbot interaction, questions inspired by the System Usability Scale (SUS) [34] were added to the survey completed after interacting with the chatbot and after seeing the group decisions. All measurements employed a seven-point bipolar Likert scale [9] (1 = strongly disagree, 2 =

[1] https://telegram.org/faq#q-what-is-telegram-what-do-i-do-here
[2] https://github.com/python-telegram-bot/
[3] https://github.com/joongishin/Consensus-building-bot

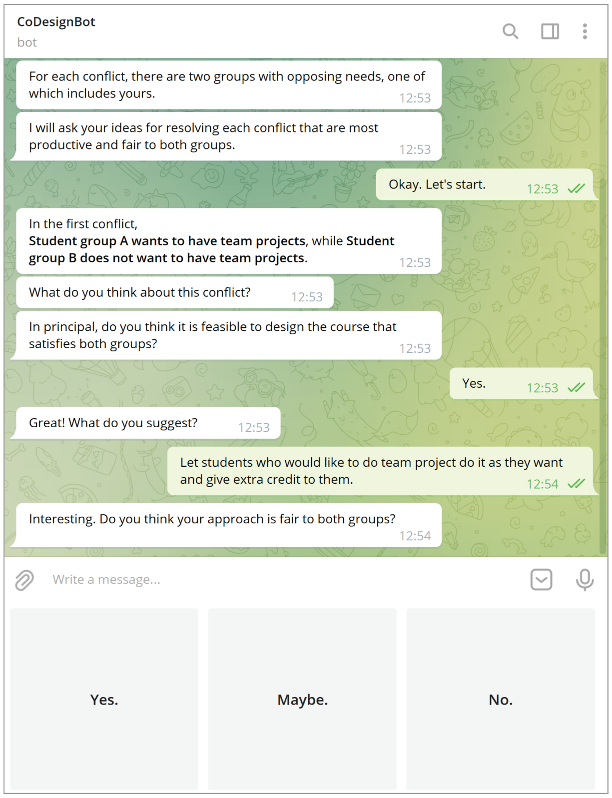

Figure 4: We implemented our chatbot using the Telegram messaging platform. The answer buttons were implemented to reduce users' burden of typing repetitive answers.

disagree, 3 = somewhat disagree, 4 = neutral or don't know, 5 = somewhat agree, 6 = agree, and 7 = strongly agree).

To examine the user responses further, we prepared a semi-structured interview protocol focused on the chatbot interaction's effect on sharing personal opinions, thinking from the others' perspective, and perceiving fairness and a sense of group effort toward shared aims. Participants' views on the benefits and limitations of this interaction as compared to face-to-face group discussion were collected as well.

4.6 Procedure

We conducted the experiment via Zoom, one participant at a time. Throughout the experiment, the participants remained anonymous and did not communicate with each other. Before day 1, we obtained the participants' consent and instructed them to be present through a desktop or laptop. They were given PDF guidance for preparing the chatbot interface themselves on their device.

On day 1, we met each participant online to introduce the purpose of the experiment as a co-design project and the general concept of co-design. To prime the participants with similar amounts of knowledge about co-design, we gave a five-minute presentation following our script and resolved any misunderstanding about the concept. During the introduction, no information was given on how they should behave – e.g., whether to give thoughtful consideration to the others' opinions or pursue shared interests. We continued by informing them about their tasks and the overall schedule of the experiment. After checking their setup for the chatbot interaction, we concluded the session with the survey collecting the participants' needs related to course activities.

On day 2, before the chatbot interaction, we informed the participants that we wished to check their understanding of the co-design project in its current state, and we presented a survey accordingly. Then, we gave instructions for interacting with the chatbot, such as to read the dialogue from the chatbot careful, that one cannot modify previous input, and to use the answer buttons (instead of typing) when presented. Then, the participants started interacting with the chatbot, without further interference from us. The chatbot led the discussion by following the conversation flow. After the participants completed the interaction, we conducted another survey and concluded the session with an interview about their experience with the chatbot.

On the last day, we told the participants that everyone had completed the discussion with the chatbot and that group decisions had been reached for each conflict, based on everyone's final suggestions at the end of the interaction. The group decisions were presented at the same time as the final survey. We randomized the order of aligned and conflicting group decisions (i.e., half of the participants saw an aligned group decision first). When the participant had completed the survey, the session concluded with the interview about the subject's opinions on each group decision and the chatbot. We recorded all participants' entire conversation with the chatbot and their interview responses. The experiment took 10, 30, and 20 minutes on each day, respectively.

5 RESULTS

All participants found the conflicts relevant to them and to reflect issues that students commonly face in their courses. Most participants did not suspect that the conflicts, suggestions from others, and group decisions were staged. Only one participant expressed suspicion, on the last day, that there were no other participants – that person had found the overall process of collecting everyone's responses through the survey and chatbot too fast and smooth. Nevertheless, that participant completed all surveys and interviews without knowing the truth.

On average, the participants interacted with the chatbot for 12.7 minutes (SD = 4.7), without needing any assistance from the research moderator. Most participants found the conversation easy to follow in general and enjoyed talking with the chatbot (see Figure 5). More specifically, they found the chatbot useful for resolving conflicts and building consensus, and felt that it had convinced them to adopt a different solution.

To further examine the influence of the chatbot on the participants' final suggestions, we reviewed all 24 discussions. The participants initially suggested voting in five of them, in three of which, on account of interacting with the chatbot, the participants changed their mind and suggested designing solutions. The participants commented that seeing the others' ideas about designing solutions had helped them recognize the possibility of satisfying both sides. In the other two cases, the participants stuck with the voting, concluding that the others' designing solutions were not practical enough.

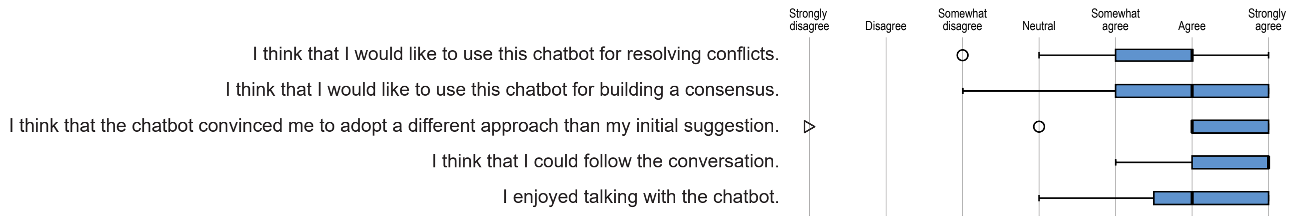

Figure 5: The results of the survey on the utility of the chatbot interaction.

In the other 19 discussions, the participants suggested designing solutions first. Here, interacting with the chatbot led the participants ultimately to pick the others' designing solutions in 15 cases. These participants commented that they found the others' solutions more practical, more productive for both sides, or similar to their own suggestions but stated more concisely. In the other four cases, the participants stood by their initial ideas, which they expected to prove more practical. Likewise, none of the participants changed their mind in favor of voting; they considered voting the less productive solution.

We performed a Wilcoxon signed-rank test (non-parametric within-subject test) to compare user responses to the survey from before and after iterating with the chatbot as well as seeing the aligned and conflicting group decisions. We report the effect size (r) of each observation using Cohen criteria of 0.1 = small effect, 0.3 = medium effect, and 0.5 = large effect. IBM SPSS statistics 26 was used, with a p-value below 0.05 indicating statistically significance.

5.1 The Chatbot's Influence on Consensus-Building

The results of the survey before and after the chatbot interaction are shown in Figure 6. A statistically significant influence of the interaction on consensus-building was visible for all items but 'I considered group benefits' and 'other participants will consider group benefits'. We assume that the concept of co-design primed the participants to aim for their shared interest and to expect the others to do the same.

As for the rest of the items, interacting with the chatbot elicited statistically significant changes in subjects' expectations of the others. The participants had a greater sense of other participants' presence in the co-design project ($Z = -2.222$, $p = 0.026$, $r = 0.454$). The post-interview revealed that seeing the others' various suggestions made them feel aware of the other participants. For instance, P1 and P5 were impressed that the others had ideas similar to theirs, and P7 noted that the interaction felt more "live" through seeing the others' suggestions in the form of quotes. Thus, they perceived others' effort and presence without having to meet in person.

The chatbot interaction significantly increased the participants' empathy with the others' suggestions ($Z = -2.310$, $p = 0.021$, $r = 0.472$) and consideration of those suggestions as potential solutions ($Z = -2.236$, $p = 0.026$, $r = 0.456$). Such perceptions appeared to influence the participants' expectations of the others; they believed that the chatbot facilitated fair contributions from all. Statistically significant increases were observed also in the participants' belief that their suggestions would be met with empathy from the others ($Z = -2.714$, $p = 0.007$, $r = 0.554$) and get regarded as potential solutions ($Z = -2.428$, $p = 0.015$, $r = 0.496$). During the interviews, participants specified that the perspective-taking and voting among the proposed solutions had led them to believe that all suggestions would be examined by everyone. P3 commented that he expected the others to behave as he did: to be convinced and to reason with the others' suggestions in the process.

The same was true for the participants' willingness to commit to the group decisions. Interacting with the chatbot had a statistically significant positive effect on that willingness ($Z = -2.333$, $p = 0.020$, $r = 0.476$) and on the level of commitment they expected from the others ($Z = -2.111$, $p = 0.035$, $r = 0.431$). The participants were confident that appropriate solutions can be proposed through the chatbot, because the interaction required them to rethink the fairness and productivity of their own proposals as well as others'. P9 said, "I did like that I could vote for any of them. I could vote for my own solution, but it was not the best one." Likewise, p12 commented that she appreciated the chatbot asking for both solutions and rationale, thus getting her to think more deeply about her initial ideas. Looking at the results, we believe that the chatbot not only caused the participants to suggest solutions that consider the whole group but also got them to sense the others' effort toward shared benefits.

5.2 Influence of Group Decisions on Consensus-Building

The participants' final suggestions represent both aligned and conflicting group decisions were provided. All participants were satisfied when seeing a group decision aligned with their personal views; they believed that fair consideration was given to all suggestions and did not express remaining curiosity (e.g., wanting to know the others' rationale for supporting the same solution). In contrast, the participants expressed disagreement with conflicting group decisions and wanted to know why the opposing side supported a particular solution.

The conflicting group decision was 'designing solutions' for two participants and 'voting' for the other 10. In both cases, the participants expressed slight disappointment and confusion during the post-interview. P1 and P3 commented that they thought the others just wanted the easiest solution, believing that voting requires less effort than designing solutions. Similarly, those participants who suggested voting deemed designing solutions was impractical when compared to voting.

Despite their dissatisfaction in the conflicting group decisions, the participants were willing to commit to aligned and conflicting group decisions both (Figure 7); the decision showed no statistically significant influence on the participants' willingness to commit

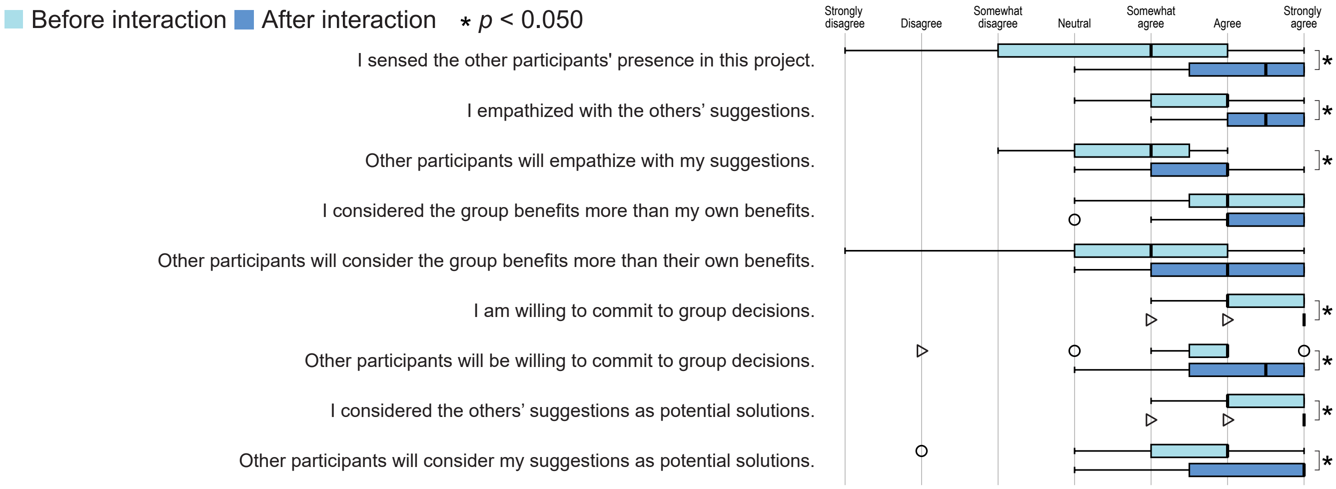

Figure 6: The survey data from before and after interaction with the chatbot. The Wilcoxon signed-rank test revealed a statistically significant influence of the chatbot on users' willingness to commit to group decisions.

to group decisions (Z = -1.841, p = 0.066, r = 0.376), perception of individual's own empathy (Z = -1.725, p = 0.084, r = 0.352), caring about the benefits for the group (Z = -1.508, p = 0.132, r = 0.308), and consideration of the others' suggestions (Z = -1.342, p = 0.180, r = 0.274). The participants reported assuming that their opinions might have been minority ones or guessing that the others saw something good in the conflicting group decisions. P3 and P12 supposed that the others reached their conclusion from different perspectives and on the basis of some reasons that they themselves did not see. P3 added that "I am willing to adapt for the better good of the group, but I am not with the statement 100 percent." Similarly, P2, with a sense of fair discussion with the chatbot, said, "However, because I had a chance to convince someone else, I am actually not dissatisfied." We believe that their striving toward common interests and their sense of fair contributions to the group's decision continued and influenced the participants to reach a compromise with conflicting group decisions.

Relative to what the participants perceived in themselves, their expectations of the others were statistically lower in the case of conflicting group decisions. The participants agreed less with the statements that the others empathized with my suggestions (Z = -2.956, p = 0.003, r = 0.603), considered my suggestions as potential solutions (Z = -2.388, p = 0.017, r = 0.487), and considered group benefits (Z = -2.584, p = 0.010, r = 0.527). The post-interview revealed that, because the subjects found their own suggestions better than the conflicting group decision, they assumed that their rationale might not have been shared properly. For instance, P10 wondered whether a different set of solutions was shown to the others. Likewise, these participants gave statistically lower scores for the chatbot's usefulness for resolving conflicts (Z = -2.333, p = 0.020, r = 0.476) and building consensus (Z = -2.060, p = 0.039, r = 0.420).

In cases of a conflicting group decision, the participants also had lower expectations for the others' willingness to commit to the group decision (Z = -2.356, p = 0.018, r = 0.481). They commented that they were uncertain how many of the others agreed or disagreed with the group decision, hence they could not confidently judge whether the others would commit to it. For instance, P11 noted that it was difficult to assess the situation without knowing how many people actually agreed with the group decision.

6 PARTICIPANT EXPERIENCE AND SUGGESTIONS

The results show that the rule-based chatbot was able to increase participants' willingness to commit to group decisions by highlighting a) the others' effort to reach shared interests and b) giving a fair chance to all participants, hence facilitating consensus-building. Proceeding from the results and self-reporting in the post-interview, we consider the implications for designing chatbots for consensus-building and discuss how chatbots would benefit co-design projects.

6.1 Implications for Design

In a first step for investigating chatbots in consensus-building, we designed the chatbot to ask questions in a specific order and respond to the participants' input. Whereas most participants found the conversation easy to follow, we did see space for improving users' engagement with the chatbot and with the others' ideas. Accordingly, we share four design implications to enhance consensus-building through a chatbot.

6.1.1 Guiding users to better promote their ideas in the dialogue. Our chatbot facilitated the asynchronous interaction and allowed the participants to take as much time as they needed for elaborating their own thoughts. Despite this benefit, the lack of immediate feedback from other group members is indeed a limitation of the asynchronous interaction. This can lower the quality of consensus-building by not identifying misunderstandings on each person's suggestions and not providing better descriptions. In response, we argue that chatbots should provide a clear instruction to eliminate any uncertainty when participants propose their ideas. For instance, P6 tended to give long answers (nearly a full paragraph) offering

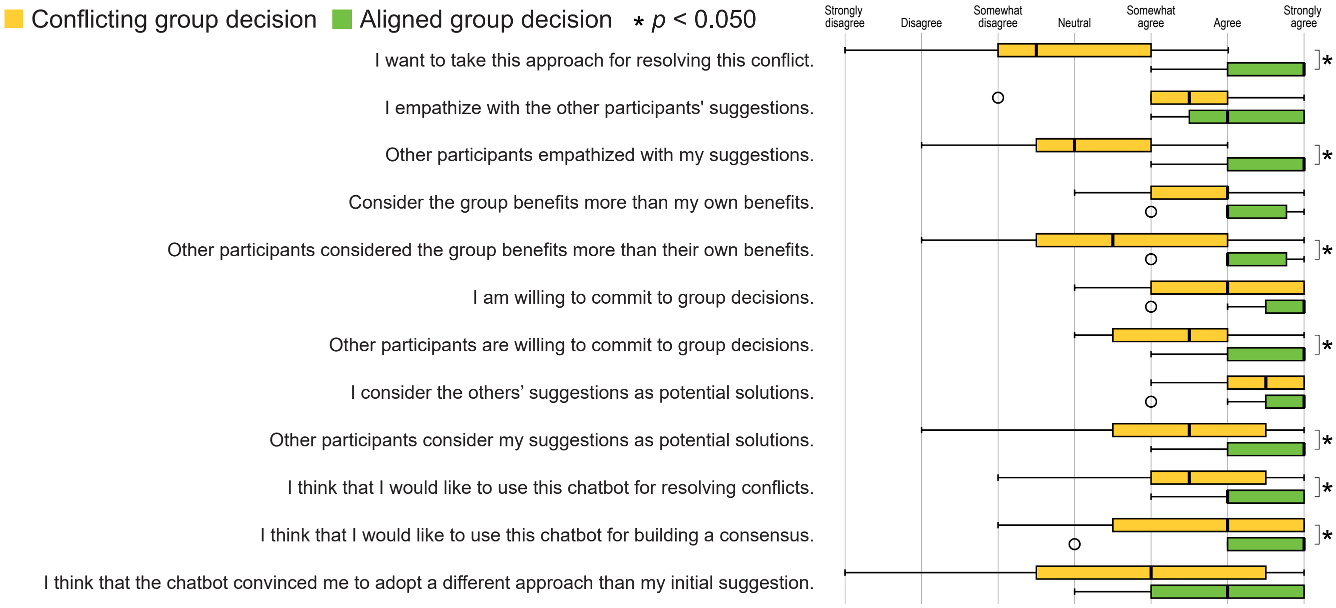

Figure 7: User responses to aligned and conflicting group decisions, collected via the survey. From our Wilcoxon signed-rank test, we observed no statistically significant difference in the participants' willingness to commit to aligned vs. conflicting group decisions. In contrast, their expectations for the others' behavior showed a significant difference.

his suggestion and detailed rationale. He stated that, because he did not know the form in which the answer should be given to the chatbot or would be shared to the others, he tried to give as much detail as possible:

> "It would have been nicer if I knew what kinds of answers I should provide. Should I tell the answer briefly or should I give reasons as well?"

P1, meanwhile, switched to shorter answers after seeing the chatbot's question about rationales and the others' responses as a single sentence. Their responses showed that users adapt to other members' writing style over time, believing that is the right format to follow. Whereas this might resolve the uncertainty in the interaction, it can harm the quality of idea exchange if the initial users describes their thoughts ineffectively (e.g., verbose). Therefore, chatbots should suggest a writing format for describing personal thoughts in the beginning of the interaction.

In another case, not knowing the conversation flow caused participants to miss opportunities to share more details. A couple of participants reported having expectation to get more chances to promote their suggestions' advantages over others'; however, the chatbot moved on to the next conflict after asking them to select one solutions. P2 commented that he would have given more details if he had known that there would be no further chances to do so. P6 likewise wanted to share more of the reasons behind specific solutions, but the opportunity never came. Accordingly, chatbots should allow participants to edit their previous input. Potentially, participants may have better thoughts after taking a break from the interaction and may want to promote their ideas in a different perspective. If participants must finalize their responses before moving on to the next dialogue, chatbots should inform this in the dialogue explicitly (e.g. signposting).

We observed a situation wherein ambiguity in the chatbot's intention left participants hesitant to answer. Three of them commented that the purpose of the perspective-taking was unclear to them since the chatbot did not explain how the participants' input would be used: P6 and P9 reported that they could not understand the purpose of expressing support for suggestions conflicted with their own, and P5 said that he wanted to skip the request since he did not want to be seen as supporting the opposite suggestions. Accordingly, we recommend resolving the issue by readying the chatbot with explanations on the purpose of its requests to users for those who demonstrate confusion.

6.1.2 Improving the Informativeness of Others' Suggestions. Presenting other group members' ideas was the key feature enabling the asynchronous consensus-building via our chatbot. The interaction made participants build on each other's viewpoints and sense group effort even without direct communication among them. User responses indicate that the interaction can improve the quality of consensus-building further by timing the moment of presenting other group members' ideas and by adjusting their similarity to the current user's ideas. In this study, the chatbot presented the others' solutions after the participants made their initial suggestions. Therefore, several participants suggested already existing ideas and later found out that there were others who had similar thoughts. From the observation, two distinctive responses were reported. The majority of the participants enjoyed that the others had similar ideas and expressed a sense of belonging, which positively influenced consensus-building.

With the current implementation, two participants viewed similar suggestions as duplicates that made the conflict resolution less efficient. They commented also that seeing the others' suggestions before offering theirs would have encouraged improving the suggestions or generating new ideas. P11 said,

> "After seeing the others' suggestions, sometimes I wanted to undo my answer because my suggestions were similar to the others, considering that this is to resolve conflicts."

In light of how people build on each other's ideas during a discussion, presenting the others' suggestions up front indeed has benefits. However, we would expect interacting with other suggestions before offering one's own thoughts to pose risks of groupthink [21] or discourage further exploration. Therefore, we suggest the chatbot to adapt its purpose of showing the others' ideas considering the phase of consensus-building. For instance, chatbots should present only the common ideas in the beginning, to elicit a sense of belonging, and present semantically different ideas later on, as inspiration to generating innovative solutions [4, 15].

Another design consideration in showing other group members' idea is information that can be collected and presented alongside their ideas. Our chatbot showed the others' suggestions as quotes, thus enhancing the sense of mutual presence. While the participants considered the quotes vivid, we observed a need for contextual information such as how confident the others were when making the point in the conversation. For instance, P4 wanted to know whether a given suggestion was an initial vs. final one and whether the person still supports the suggestion as stated. To improve users' empathy for and understanding of other group members' viewpoints, we recommend having means for users to express their attachment to their ideas in the dialogue (e.g. emoticon). Potentially, chatbots may deliver such information by re-creating the dialogue in which other group members shared their thoughts, allowing participants to better understand the others' thought process.

6.1.3 Moderating Conflicts on Group Decisions. This study investigated the influence of interacting with the chatbot over only a single cycle. The chatbot's ability to increase participants' willingness to commit to group decisions points to a need for further interaction with the chatbot, related to convincing the others. In response, the participants wanted review the details about the group decisions that conflicted with personal interests. The most commonly requested information was the number of participants who (dis)agreed with the ultimate decisions – participants were concerned that there might be many others who disagreed with the group decisions, and P4 added that knowing how many people voted for each approach would have helped him understand how his group is navigating at the moment. Also, the participants wanted to review other group members' reasons for agreeing with the conflicting group decisions. They wanted to examine possible misunderstandings or a perspective that they might have missed. The most representative expression of this need came from P12: "If I cannot deny it, at least I want to know why." On the contrary, if a large number of participants are invited through chatbots, reviewing all members' opinions would be time-consuming and identifying major opinions would be difficult. We argue, in response, that chatbots should provide an efficient way of examining the complexity behind group decisions. For instance, system could use NLP methods such as semantic clustering to identify the most representative opinions to focus on or adaptively present opinions contrasting against the individuals' previous viewpoints.

In addition to reviewing other group members' input, the participants wanted to express their (dis)agreement with the group decisions, regardless of their commitment. They expected such information to prompt further discussion or reconsideration on suggestions receiving fewer votes. Similar to how the chatbot collected everyone's suggestions and arrived at the group decision accordingly, the chatbot could meet with the individuals again, to assess their agreement and propose multiple rounds of discussion. In addition, informing users that further discussion will follow might accentuate the consensus-building's fairness further.

6.1.4 Moderating User Inputs to Fight Malevolent Posting. Throughout the study, the participants gave thoughtful answers to the chatbot and perceived the others' suggestions to be reasonable. However, in anonymous settings, it could happen that not all participants give sincere suggestions. Potentially, some participants might spam their own ideas regardless of what the chatbot asks or give very brief responses to complete the interaction as quickly as possible. In the absence of any filtering system, we would expect such responses to get shared with all participants and disrupt their focus on actual conflicts. Work on NLP has studied detection of unsuitable or malicious user content extensively, covering use cases such as spam [42], fake reviews [43] and hate speech [11]. The approaches range from simple rule-based systems to advanced methods of machine learning. While an automated system may not detect malevolent posting perfectly, it could reduce its negative impacts by pre-filtering or by informing a human moderator about possible issues.

6.2 Asynchrony and Anonymity in Co-Design

Our study showed that a chatbot can facilitate consensus-building without having the participants communicate with each other, hence enabling asynchronous interaction. The results suggest that chatbots can asynchronously conduct co-design activities [8] wherein stakeholders contribute in their own time and space. We argue that this is consistent with the core value of co-design: involving real end-users instead of only a group of representatives. Without set schedules, any users who have access to the chatbot could join not only to merely share their opinions [39] but also to participate in constructive co-design activities moderated by chatbots. In contrast to group activities with a single moderator and multiple stakeholders, chatbots show promise for meeting with individual users to enable interaction with the others' opinions at their own comfortable pace.

We should stress that enabling users to participate in their own environment is another attention-worthy benefit. Designers and researchers have employed diverse methods to collect sticky information [58] in users' contexts, such as ethnography, contextual inquiry, diary and camera studies [40], yet most co-design activities still bring users into one place, outside their own environment. Potentially, chatbots can moderate co-design activities in users' spaces, where sharing contextual knowledge may be easier. That said, we believe there is a trade-off to the benefits of asynchronous co-design, in the duration of co-design projects. In most co-design projects, stakeholders gather and contribute altogether in, all told, 1-2 hours of co-design activities, then can easily move on to the subsequent phases. In contrast, asynchronous co-design with a chatbot would extend each phase since stakeholders need to wait until the rest of the stakeholders finish interacting with the chatbot. Managing the deadline for each phase and overall progress constitutes another design challenge.

Unlike in-person co-design activities, our chatbot setting hid the participants' identity and showed that everyone's suggestions were shared anonymously. The interactions were always between each participant and the others' ideas, without any social interaction taking place. Whereas socio-emotional factors are considered to be essential for consensus-building [20, 24, 55], the participants reported benefits from having no social interaction, such as having an objective view of the conflicts and potential solutions. For instance, P8 specified that interaction via the chatbot removed any distraction caused by seeing the others' emotions and that they could focus on the facts:

> "I think it is good that you do not see the other people, so there is no issue of ego and no need to be afraid of being misjudged when suggesting ideas."

Similarly, P7 and P9 commented that not seeing the others helped her evaluate their suggestions without worrying about harming their ego. Such anonymity could be further explored in a co-design context, to encourage shy participants or elicit bold ideas that, while perhaps less practical, could inspire others.

However, combining anonymity with asynchrony presents potential risks, such as stakeholders with higher authority navigating co-design. For instance, in our study context, lecturers could prepare solutions such that the chatbot presents them as if coming from other students involved in the project. Since the responses are not labeled and the chatbot does not provide direct communication among students, it could be difficult for them to spot such manipulation. In the case of co-design in a small group (e.g., under 10 stakeholders), it might be easier to cross-check stakeholders' input in the event of doubt, but in co-design projects with a large number of stakeholders, expressing doubts would be difficult from the outset – stakeholders could presume their opinions to be minority ones. Further investigation would be required for designing a fail-safe or reliable moderation.

7 CONCLUSION AND FUTURE WORK

With this paper, we have introduced the idea that a chatbot could moderate consensus-building in co-design by engaging in conflict resolution with the individual stakeholders. Our study shows that presenting the others' responses and encouraging perspective-taking in the conversation can raise users' willingness to commit to group decisions. Even when disagreeing with them, the participants were willing to accept the group decisions for the good of everyone. At the same time, we observed a need for iteratively sharing each person's rationale and expressing one's (dis)agreement with the group decisions. The chatbot as a moderator brings the benefits of asynchronicity and anonymity into co-design.

To study conflicts with relevance for each participant, we generated the aligned and conflicting opinions ourselves. We expect the future to enable replacing such effort with NLP-based estimation of the semantic similarity between the current user's input and other group members' ideas, suggested earlier. Potentially, NLP could also be combined with a predefined conversation flow [38] and adaptively guide a set of consensus-building activities.

Building on our results on consensus-building, future work should include investigation of chatbot interaction over the full cycle of co-design projects, whereas our study was limited to chatbot interaction in the earlier stages of design, when stakeholders' ideas remain at the conceptual level. Exploring how chatbots lead asynchronous implementation and testing of solutions is another fascinating avenue for extending co-design. Potentially, other forms of ideation (e.g., drawings and video) could be fruitfully integrated with chatbots, in light of how co-design promotes multi-modal communication to enhance stakeholders' idea exchange. Another limitation of our study is that the interaction was in an experiment setting. Since the group decisions in our study did not have a critical impact on the participants, they might have been more accepting of the conflicting group decisions than in a real-world setting. Further investigation is required for understanding how chatbots can mediate conflicts that could severely harm a portion of the group. Nevertheless, our study showed that the interaction increased the consideration for other members' efforts and of their suggestions, which are essential for consensus-building.

ACKNOWLEDGMENTS

The research was supported by the Aalto University Department of Communications and Networking, the Finnish Center for Artificial Intelligence (FCAI), Academy of Finland (grants 'Human Automata' and 'BAD'), and a fellowship within the IFI programme of the German Academic Exchange Service (DAAD).

REFERENCES

[1] 2005. *Concise Interactions and Effective Management of Shared Design Spaces: Moving Beyond Strategic Collaboration Towards Co-Design.* International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. Volume 5a: 17th International Conference on Design Theory and Methodology. https://doi.org/10.1115/DETC2005-85381 arXiv:https://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2005/4742Xa/231/2643348/231_1.pdf

[2] Pernille Viktoria Kathja Andersen and Wafa Said Mosleh. 2021. Conflicts in co-design: engaging with tangible artefacts in multi-stakeholder collaboration. *CoDesign* 17, 4 (2021), 473–492. https://doi.org/10.1080/15710882.2020.1740279 arXiv:https://doi.org/10.1080/15710882.2020.1740279

[3] Pernille Viktoria Kathja Andersen and Wafa Said Mosleh. 2021. Conflicts in co-design: engaging with tangible artefacts in multi-stakeholder collaboration. *CoDesign* 17, 4 (2021), 473–492. https://doi.org/10.1080/15710882.2020.1740279 arXiv:https://doi.org/10.1080/15710882.2020.1740279

[4] Suyun Sandra Bae, Oh-Hyun Kwon, Senthil Chandrasegaran, and Kwan-Liu Ma. 2020. *Spinneret: Aiding Creative Ideation through Non-Obvious Concept Associations.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376746

[5] Hamish Beattie, Daniel K. Brown, and Morten Gjerde. 2017. Generating Consensus: A Framework for Fictional Inquiry in Participatory City Gaming. In *Serious Games*, Mariano Alcañiz, Stefan Göbel, Minhua Ma, Manuel Fradinho Oliveira, Jannicke Baalsrud Hauge, and Tim Marsh (Eds.). Springer International Publishing, Cham, 126–137.

[6] Robert Briggs, Gwendolyn Kolfschoten, and Gert-Jan de Vreede. 2005. Toward a Theoretical Model of Consensus Building. 12.

[7] Peter T Coleman, Morton Deutsch, and Eric C Marcus. 2014. *The handbook of conflict resolution: Theory and practice.* John Wiley & Sons.

[8] Aaron Davis, Wallace Niki, Joseph Langley, and Gwilt Ian. 2021. Low-contact co-design: considering more flexible spatiotemporal models for the co-design workshop. *Strategic Design Research Journal* 14, 1 (2021).

[9] John Dawes. 2008. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International journal of market research* 50, 1 (2008), 61–104.

[10] Larry Dressler. 2006. *Consensus Through Conversations: How to Achieve High-Commitment Decisions*. Berrett-Koehler Publishers.

[11] Paula Fortuna and Sérgio Nunes. 2018. A Survey on Automatic Detection of Hate Speech in Text. *ACM Comput. Surv.* 51, 4, Article 85 (jul 2018), 30 pages. https://doi.org/10.1145/3232676

[12] Pirjo Friedrich. 2013. *Web-based co-design: Social media tools to enhance user-centred design and innovation processes: Dissertation*. Ph. D. Dissertation. Aalto University, Finland.

[13] Elisa Giaccardi, Pedro Paredes, Paloma Díaz, and Diego Alvarado. 2012. Embodied Narratives: A Performative Co-Design Technique. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) *(DIS '12)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/2317956.2317958

[14] Yoshiko Goda, Masanori Yamada, Hideya Matsukawa, Kojiro Hata, and Seisuke Yasunami. 2014. Conversation with a Chatbot before an Online EFL Group Discussion and the Effects on Critical Thinking. *The Journal of Information and Systems in Education* 13, 1 (2014), 1–7. https://doi.org/10.12937/ejsise.13.1

[15] Milene Gonçalves, Carlos Cardoso, and Petra Badke-Schaub. 2013. Inspiration peak: exploring the semantic distance between design problem and textual inspirational stimuli. *International Journal of Design Creativity and Innovation* 1, 4 (2013), 215–232. https://doi.org/10.1080/21650349.2013.799309 arXiv:https://doi.org/10.1080/21650349.2013.799309

[16] P. Grünbacher and R.O. Briggs. 2001. Surfacing tacit knowledge in requirements negotiation: experiences using EasyWinWin. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. 8 pp.–. https://doi.org/10.1109/HICSS.2001.926243

[17] Rafik Hadfi, Jawad Haqbeen, Sofia Sahab, and Takayuki Ito. 2021. Argumentative conversational agents for online discussions. *Journal of Systems Science and Systems Engineering* 30, 4 (2021), 450–464. https://doi.org/10.1007/s11518-021-5497-1

[18] Kim Halskov and Peter Dalsgård. 2006. Inspiration Card Workshops. In *Proceedings of the 6th Conference on Designing Interactive Systems* (University Park, PA, USA) *(DIS '06)*. Association for Computing Machinery, New York, NY, USA, 2–11. https://doi.org/10.1145/1142405.1142409

[19] Jawad Haqbeen, Takayuki Ito, Rafik Hadfi, Tomohiro Nishida, Zoia Sahab, Sofia Sahab, Shafiq Roghmal, and Ramin Amiryar. 2020. Promoting discussion with AI-based facilitation: Urban dialogue with Kabul city. In *Proceedings of the 8th ACM Collective Intelligence, ACM Collective Intelligence Conference Series, Boston (Virtual Conference), South Padre Island, TX, USA*, Vol. 18.

[20] Ning He and Yang Liu. 2019. An Application of Support States from Speech Emotions in Consensus Building. In *Genetic and Evolutionary Computing*, Jeng-Shyang Pan, Jerry Chun-Wei Lin, Bixia Sui, and Shih-Pang Tseng (Eds.). Springer Singapore, Singapore, 321–329.

[21] Irving L Janis. 2008. Groupthink. *IEEE Engineering Management Review* 36, 1 (2008), 36.

[22] Sanna Järvelä and Päivi Häkkinen. 2002. Web-based Cases in Teaching and Learning – the Quality of Discussions and a Stage of Perspective Taking in Asynchronous Communication. *Interactive Learning Environments* 10, 1 (2002), 1–22. https://doi.org/10.1076/ilee.10.1.1.3613 arXiv:https://doi.org/10.1076/ilee.10.1.1.3613

[23] Anu Kankainen, Kirsikka Vaajakallio, Vesa Kantola, and Tuuli Mattelmäki. 2012. Storytelling Group – a co-design method for service design. *Behaviour & Information Technology* 31, 3 (2012), 221–230. https://doi.org/10.1080/0144929X.2011.563794 arXiv:https://doi.org/10.1080/0144929X.2011.563794

[24] Yasuhiro Katagiri, Katsuya Takanashi, Masato Ishizaki, Yasuharu Den, and Mika Enomoto. 2013. Concern Alignment and Trust in Consensus-building Dialogues. *Procedia - Social and Behavioral Sciences* 97 (2013), 422–428. https://doi.org/10.1016/j.sbspro.2013.10.254 The 9th International Conference on Cognitive Science.

[25] Nick Kelly, Natalie Wright, Les Dawes, Jeremy Kerr, and Amanda Robertson. 2019. *Australian Journal of Teacher Education (Online)* 44, 7 (2019), 84–107. https://search.informit.org/doi/10.3316/informit.637458284530867

[26] Alison Kennedy, Catherine Cosgrave, Joanna Macdonald, Kate Gunn, Timo Dietrich, and Susan Brumby. 2021. Translating Co-Design from Face-to-Face to Online: An Australian Primary Producer Project Conducted during COVID-19. *International Journal of Environmental Research and Public Health* 18, 8 (2021). https://doi.org/10.3390/ijerph18084147

[27] Finn Kensing and Andreas Munk-Madsen. 1993. PD: Structure in the Toolbox. *Commun. ACM* 36, 6 (1993), 78–85.

[28] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joonhwan Lee. 2020. *Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376785

[29] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator Chatbot for Deliberative Discussion: Effects of Discussion Structure and Discussant Facilitation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 87 (apr 2021), 26 pages. https://doi.org/10.1145/3449161

[30] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300316

[31] Scott R. Klemmer, Mark W. Newman, Ryan Farrell, Mark Bilezikjian, and James A. Landay. 2001. The Designers' Outpost: A Tangible Interface for Collaborative Web Site. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology* (Orlando, Florida) *(UIST '01)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/502348.502350

[32] Sung-Chul Lee, Jaeyoon Song, Eun-Young Ko, Seongho Park, Jihee Kim, and Juho Kim. 2020. *SolutionChat: Real-Time Moderator Support for Chat-Based Structured Discussion*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376609

[33] Terry TF Leung and Barry CL Lam. 2019. Building consensus on user participation in social work: A conversation analysis. *Journal of Social Work* 19, 1 (2019), 20–40. https://doi.org/10.1177/1468017318757357 arXiv:https://doi.org/10.1177/1468017318757357

[34] James R. Lewis. 2018. The System Usability Scale: Past, Present, and Future. *International Journal of Human–Computer Interaction* 34, 7 (2018), 577–590. https://doi.org/10.1080/10447318.2018.1455307 arXiv:https://doi.org/10.1080/10447318.2018.1455307

[35] Weichen Liu, Sijia Xiao, Jacob T. Browne, Ming Yang, and Steven P. Dow. 2018. ConsensUs: Supporting Multi-Criteria Group Decisions by Visualizing Points of Disagreement. *Trans. Soc. Comput.* 1, 1, Article 4 (jan 2018), 26 pages. https://doi.org/10.1145/3159649

[36] Andrés Lucero. 2015. Using Affinity Diagrams to Evaluate Interactive Prototypes. In *Human-Computer Interaction – INTERACT 2015*, Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler (Eds.). Springer International Publishing, Cham, 231–248.

[37] Andrés Lucero, Kirsikka Vaajakallio, and Peter Dalsgaard. 2012. The dialogue-labs method: process, space and materials as structuring elements to spark dialogue in co-design events. *CoDesign* 8, 1 (2012), 1–23. https://doi.org/10.1080/15710882.2011.609888 arXiv:https://doi.org/10.1080/15710882.2011.609888

[38] Wookjae Maeng and Joonhwan Lee. 2021. Designing a Chatbot for Survivors of Sexual Violence: Exploratory Study for Hybrid Approach Combining Rule-Based Chatbot and ML-Based Chatbot. In *Asian CHI Symposium 2021* (Yokohama, Japan) *(Asian CHI Symposium 2021)*. Association for Computing Machinery, New York, NY, USA, 160–166. https://doi.org/10.1145/3429360.3468203

[39] M Lynne Markus and Ji-Ye Mao. 2004. Participation in development and implementation-updating an old, tired concept for today's IS contexts. *Journal of the Association for Information systems* 5, 11 (2004), 14.

[40] B. Martin, B. Hanington, and B.M. Hanington. 2012. *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Rockport Publishers. https://books.google.fi/books?id=uZ8uzWAcdxEC

[41] Ali Mazalek, Claudia Winegarden, Tristan Al-Haddad, Susan J. Robinson, and Chih-Sung Wu. 2009. Architales: Physical/Digital Co-Design of an Interactive Story Table. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction* (Cambridge, United Kingdom) *(TEI '09)*. Association for Computing Machinery, New York, NY, USA, 241–248. https://doi.org/10.1145/1517664.1517716

[42] Vangelis Metsis, Ion Androutsopoulos, and Georgios Paliouras. 2006. Spam Filtering with Naive Bayes - Which Naive Bayes?. In *CEAS 2006 - The Third Conference on Email and Anti-Spam, July 27-28, 2006, Mountain View, California, USA*. http://www.ceas.cc/2006/listabs.html#15.pdf

[43] Rami Mohawesh, Shuxiang Xu, Son N. Tran, Robert Ollington, Matthew Springer, Yaser Jararweh, and Sumbal Maqsood. 2021. Fake Reviews Detection: A Survey. *IEEE Access* 9 (2021), 65771–65802. https://doi.org/10.1109/ACCESS.2021.3075573

[44] Pirjo Näkki and Maria Antikainen. 2008. Online tools for co-design: User involvement through the innovation process. *New Approaches to Requirements Elicitation* 96 (2008).

[45] Jaya Narain, Tina Quach, Monique Davey, Hae Won Park, Cynthia Breazeal, and Rosalind Picard. 2020. Promoting Wellbeing with Sunny, a Chatbot That Facilitates Positive Messages within Social Groups. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3383062

[46] SangAh Park, Yoon Young Lee, Soobin Cho, Minjoon Kim, and Joongseek Lee. 2021. *"Knock Knock, Here Is an Answer from Next Door": Designing a Knowledge Sharing Chatbot to Connect Residents: Community Chatbot Design Case Study*. Association for Computing Machinery, New York, NY, USA, 144–148. https://doi.org/10.1145/3462204.3481738

[47] Frank Piller, Petra Schubert, Michael Koch, and Kathrin Möslein. 2017. Overcoming Mass Confusion: Collaborative Customer Co-Design in Online Communities. *Journal of Computer-Mediated Communication* 10, 4 (07 2017). https://doi.org/10.1111/j.1083-6101.2005.tb00271.x JCMC1042.

[48] Antti Pirinen et al. 2016. The barriers and enablers of co-design for services. *International Journal of Design* 10, 3 (2016), 27–42.

[49] Jungwook Rhim, Minji Kwak, Yeaeun Gong, and Gahgene Gweon. 2022. Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. *Computers in Human Behavior* 126 (2022), 107034. https://doi.org/10.1016/j.chb.2021.107034

[50] Toni Robertson and Jesper Simonsen. 2012. Challenges and Opportunities in Contemporary Participatory Design. *Design Issues* 28, 3 (07 2012), 3–9. https://doi.org/10.1162/DESI_a_00157 arXiv:https://direct.mit.edu/desi/article-pdf/28/3/3/1715026/desi_a_00157.pdf

[51] Elizabeth B.-N. Sanders and Pieter Jan Stappers. 2008. Co-creation and the new landscapes of design. *CoDesign* 4, 1 (2008), 5–18. https://doi.org/10.1080/15710880701875068 arXiv:https://doi.org/10.1080/15710880701875068

[52] Jihyeong Son, Amrut Sadachar, Srikant Manchiraju, Ann Marie Fiore, and Linda S Niehm. 2012. Consumer adoption of online collaborative customer co-design. *Journal of Research in Interactive Marketing* (2012). https://doi.org/10.1108/17505931211274660

[53] Ryota Sugino, Satoshi Mizoguchi, Koji Kimita, Keiichi Muramatsu, Tatsunori Matsui, and Yoshiki Shimomura. 2016. A Method for Consensus Building Between Teachers and Learners in Higher Education Through Co-design Process. In *Human Interface and the Management of Information: Applications and Services*, Sakae Yamamoto (Ed.). Springer International Publishing, Cham, 197–208.

[54] Lawrence E Susskind, Sarah McKearnen, and Jennifer Thomas-Lamar. 1999. *The consensus building handbook a comprehensive guide to reaching agreement*. SAGE, Thousand Oaks, Calif. ;.

[55] Paul Thagard and Fred W Kroon. 2006. Emotional consensus in group decision making. *Mind & Society* 5, 1 (2006), 85–104. https://doi.org/10.1007/s11299-006-0011-5

[56] Muneyuki Unehara, Midori Saitou, Koichi Yamada, and Izumi Suzuki. 2017. Design Support System with Consensus Building of Multiple Participants by Interactive Evolutionary Computation. In *the 18th International Symposium on Advanced Intelligent Systems (ISIS2017)*. 433–441.

[57] Froukje Sleeswijk Visser, Pieter Jan Stappers, Remko van der Lugt, and Elizabeth B-N Sanders. 2005. Contextmapping: experiences from practice. *CoDesign* 1, 2 (2005), 119–149. https://doi.org/10.1080/15710880500135987 arXiv:https://doi.org/10.1080/15710880500135987

[58] Eric von Hippel. 2001. User toolkits for innovation. *Journal of Product Innovation Management* 18, 4 (2001), 247–257. https://doi.org/10.1111/1540-5885.1840247 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/1540-5885.1840247

[59] Greg Walsh, Craig Donahue, and Zachary Pease. 2016. Inclusive Co-Design within a Three-Dimensional Game Environment. In *Proceedings of the The 15th International Conference on Interaction Design and Children* (Manchester, United Kingdom) *(IDC '16)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/2930674.2930721

[60] Greg Walsh and Eric Wronsky. 2019. AI + Co-Design: Developing a Novel Computer-Supported Approach to Inclusive Design. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 408–412. https://doi.org/10.1145/3311957.3359456

[61] Thiemo Wambsganss, Tobias Kueng, Matthias Soellner, and Jan Marco Leimeister. 2021. ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 683, 13 pages. https://doi.org/10.1145/3411764.3445781

[62] Thiemo Wambsganss, Rainer Winkler, Pascale Schmid, and Matthias Söllner. 2020. Unleashing the Potential of Conversational Agents for Course Evaluations: Empirical Insights from a Comparison with Web Surveys. In *Proceedings of the 28th European Conference on Information Systems (ECIS)*.

[63] Bo Xie, Allison Druin, Jerry Fails, Sheri Massey, Evan Golub, Sonia Franckel, and Kiki Schneider. 2012. Connecting generations: developing co-design methods for older adults and children. *Behaviour & Information Technology* 31, 4 (2012), 413–423. https://doi.org/10.1080/01449291003793793 arXiv:https://doi.org/10.1080/01449291003793793

[64] Roshanak Zilouchian Moghaddam, Brian P. Bailey, and Christina Poon. 2011. IdeaTracker: An Interactive Visualization Supporting Collaboration and Consensus Building in Online Interface Design Discussions. In *Human-Computer Interaction – INTERACT 2011*, Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 259–276.