Supporting Collaborative Sequencing of Small Groups through Visual Awareness

TAE SOO KIM, School of Computing, KAIST, Republic of Korea
NITESH GOYAL, Jigsaw, Google, USA
JEONGYEON KIM, School of Computing, KAIST, Republic of Korea
JUHO KIM, School of Computing, KAIST, Republic of Korea
SUNGSOO RAY HONG, Information Sciences and Technology, George Mason University, USA

Collaborative Sequencing (CoSeq) is the process by which a group collaboratively constructs a sequence. CoSeq is ubiquitous, occurring across diverse situations like trip planning, course scheduling, or book writing. Building a consensus on a sequence is desirable to groups. However, accomplishing this requires groups to dedicate significant effort to comprehensively discuss preferences and resolve conflicts. Furthermore, as numerous decisions must be assessed to construct a sequence, this challenge can be exacerbated in CoSeq. However, little research has aimed to effectively support consensus building in CoSeq. As a first step to systematically understand and support consensus building in CoSeq, we conducted a formative study to gain insights into how visual awareness may facilitate the holistic recognition of preferences and the resolution of conflicts within a group. From the study, we identified design requirements to support consensus building and designed a novel visual awareness technique for CoSeq. We instantiated this design in a collaborative travel itinerary planning system, Twine, and conducted a summative study to evaluate its effects. We found that visual awareness could decrease the effort of communicating preferences by 21%, and participants' comments suggest that it also encouraged group members to behave more cooperatively when building a consensus. We discuss future research directions to further explore the needs and challenges in this unique context and to advance the development of support for CoSeq tasks.

CCS Concepts: • **Human-centered computing** → **Social navigation**; **Computer supported cooperative work**; • **Information systems** → *Collaborative and social computing systems and tools.*

Additional Key Words and Phrases: Collaborative Sequencing; Visual Awareness; Group Awareness; Consensus; Small Group; Group Communication

ACM Reference Format:
Tae Soo Kim, Nitesh Goyal, Jeongyeon Kim, Juho Kim, and Sungsoo Ray Hong. 2021. Supporting Collaborative Sequencing of Small Groups through Visual Awareness. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 176 (April 2021), 29 pages. https://doi.org/10.1145/3449250

1 INTRODUCTION

Collaborative Sequencing (CoSeq) is a task in which a group of members collaboratively constructs a sequence by selecting items from a set of possible alternatives and arranging these items into a

Authors' addresses: Tae Soo Kim, School of Computing, KAIST, Daejeon, Republic of Korea, taesoo.kim@kaist.ac.kr; Nitesh Goyal, Jigsaw, Google, New York, USA, teshg@google.com; Jeongyeon Kim, School of Computing, KAIST, Daejeon, Republic of Korea, imurs34@kaist.ac.kr; Juho Kim, School of Computing, KAIST, Daejeon, Republic of Korea, juhokim@kaist.ac.kr; Sungsoo Ray Hong, Information Sciences and Technology, George Mason University, Virginia, USA, shong31@gmu.edu.

2573-0142/2021/4-ART176 $15.00
https://doi.org/10.1145/3449250

particular order. People commonly engage in such tasks in both casual and formal contexts: when planning a trip with friends [48], determining a curriculum for coursework [64], structuring sections of a book, and scheduling the writing process [50]. It is desirable for group members to build a consensus regarding a sequence in scenarios where the members are collectively responsible for and affected by that sequence [10]. This is because, when compared to other group decision-making processes like majority vote, consensus building has been shown to lead to higher satisfaction regarding the task and interpersonal relationships among group members [21]. Consequently, for example, it might be more preferable for a group of tourists to reach a consensus on the sequence of attractions they will visit during their trip, than to decide by voting.

Despite the benefits of consensus building, reaching a decision based on all of the group members' consensus is challenging. For instance, members must expend significant time and communication effort to express their opinions and to develop a comprehensive awareness of the group's preferences [10]. Additionally, factors such as individual biases [60] and unequal participation in discussions [33] may further hinder the ability of groups to reach an effective consensus.

Group decision-making and consensus building have been core research topics in CSCW [6, 9, 30, 39]. For example, substantial research has been dedicated into understanding and supporting design discussions [69], criteria-based decision-making [43], and collaborative analysis [35]. However, despite the ubiquity of CoSeq, it has received relatively less attention within the CSCW and Human-Computer Interaction (HCI) communities, leaving the challenges inherent in consensus building unaddressed in CoSeq tasks. Therefore, further exploration into this unique context is required to effectively support this type of task.

The overarching goal of this work is to determine a design that can effectively facilitate consensus building in CoSeq. Prior work has demonstrated the benefit of visual awareness in decision-making contexts. Visual awareness denotes visual support which allows the user to maintain awareness of task-related elements such as group members' opinions and actions, and the overall state of the process. In several tasks, visual awareness of members' preferences in groups has been demonstrated to decrease communication effort [22] and advance discussions [30]. Inspired by these, we aimed to gain a deeper understanding of how to effectively facilitate group awareness in CoSeq. In our formative study, we generated four notable prototypes based on prior work of collaborative information seeking [29], information visualization [4], and graph theory [3, 7, 19]. Using these prototypes, we conducted a qualitative study with four users to elicit aspects that are useful for gaining awareness of preferences within a group. We boiled down these aspects into three design requirements that aim to support groups' consensus building through visual awareness. In particular, visual awareness should help groups assess the overall level of agreement, focus on the differences between members' selection and ordering of items, and guide them to determine specific actions that they could take to reach a consensus.

To evaluate how such visual awareness affects group communication and consensus building in CoSeq, we designed a collaborative travel itinerary planning system, Twine. Using Twine, we conducted a within-subjects controlled study with 45 participants—15 teams of three travelers each. Findings from our study indicate that the participants perceived the consensus building process in CoSeq to be more efficient and effective with visual awareness than without it. Regarding efficiency, we investigated the effort and time required to reach a consensus, and found that visual awareness reduced participants' effort in expressing opinions by 21% and that of inquiring about members' opinions by 22%. Regarding effectiveness, we investigated participants' perceptions regarding both the overall satisfaction of their group members during the consensus building process as well as individual satisfaction towards the final outcome. Our results showed that, while the perceived group satisfaction increased significantly with visual awareness, individual satisfaction did not. A qualitative analysis of participants' responses suggests our visual awareness technique can encourage group members to be more receptive to others' opinions, but can also potentially pressure them to conform to others' opinions. From the study findings, we also identified design implications to further facilitate and enhance groups' consensus building processes in diverse CoSeq scenarios.

This work offers the following contributions:

- **Design Requirements for a CoSeq Task - Formative Study**: Through our formative study, we characterize common challenges and needs in CoSeq and determine a set of three design requirements that could facilitate successful consensus building.
- **Design of a Research Prototype - Twine** : Informed by the design requirements, we present a novel technique that leverages visual awareness to facilitate group communication and consensus building in CoSeq.
- **Experimental Results about Efficiency and Effectiveness of Twine - Summative Study**: We present empirical evidence that shows how our design can facilitate CoSeq in terms of task efficiency and effectiveness through our collection of behavioral and attitudinal measures.

2 BACKGROUND AND RELATED WORK

We review related work in four main areas: (1) systems that support sequencing, mostly for individual users, (2) consensus building theory, (3) system-supported consensus building, and (4) techniques for sequence comparison.

2.1 Approaches to Support Sequencing

In this work, we define sequences as directed, acyclic paths of nodes (i.e., selected items). The nodes possess quantitative and/or qualitative attributes, and are interrelated such that pairs of consequent nodes are connected by edges which are weighted by quantitative and/or qualitative values. For example, in the case of book structuring, each chapter has a quantitative page length and qualitative content, and consequent chapters have a qualitative logical progression. Additionally, as a whole, sequences may be constrained (e.g., limiting the total page length of a book). Due to the number of decisions, alternatives and related factors that need to be assessed, effectively constructing a sequence can be challenging. Thus, many researchers have aimed to design systems that facilitate sequence construction—i.e., sequencing.

Recommender systems, which often leverage large-scale data, have been a common approach [11, 13, 31, 55, 58]. *Xnavi* [48] assists a traveler's itinerary planning process by recommending common subsequences of activities which are extracted from past tourists' driving histories. To provide advisors with action plan recommendations for their students, *EventAction* [14] exploits event sequences in the academic records of past students. In the space of storyboarding, Tharatipyakul et al. [59] designed a system which allows the user to sequence and color-code frames to visually compare variations of a storyboard. Alternatively, *Mobi* [66] and *Crowdcierge* [52] used crowdsourcing to harness human computation to generate travel itineraries for requesters. These systems, however, only address the sequencing tasks of individuals. In the case of the crowdsourcing systems, they supported collaborating crowdworkers but their sequence construction is based on the requirements and preferences of a single requester. Therefore, they were not designed to facilitate the decision-making process through which several individuals may discuss opinions to construct and reach an agreement on a sequence. Most related to CoSeq tasks, Cobi [37] facilitates collaborative conference scheduling through communitysourcing and constraint-solving support. However, the system focuses on the task of allocating papers into sessions and provides no support for the sequencing of papers within sessions such that they follow a logical progression. Recently, Kim et al. [38] provided preliminary observations on the effect of visual support in CoSeq. We extend this work by providing in-depth insights on the design of visual awareness for CoSeq tasks and implications for future design.

2.2 Theoretical Constructs of Group Consensus Building

CoSeq tasks involve group members jointly constructing a sequence based on their individual opinions (e.g., preferences, knowledge or expertise) regarding the multiple qualitative and quantitative attributes of nodes and edges in the sequence. Due to the complex relationships between the members' opinions and the multi-criteria nature of sequences, CoSeq tasks are problems with no correct answers—categorized as group tasks of the "decision-making task" type in McGrath's circumplex model [45]. Due to the lack of a clear solution, these types of tasks can benefit from group members building a consensus to reach a decision. Consensus building increases stakeholders willingness to commit to a proposal [8]—in the case of CoSeq, a sequence—by resolving present disagreements or conflicts to produce more high-quality and acceptable outcomes [57]. Due to the merits of consensus building, Briggs et al. [8] introduced a general process model to guide the application of this process in diverse group tasks. The model consists of four main steps: (1) a member makes a proposal, (2) the group evaluates their willingness to commit to it, (3) the group identifies conflicts, and (4) lastly, they resolve the conflicts.

While consensus building is applicable to CoSeq tasks, the suitability of the process depends on the context of the task. Firstly, the process is desirable when members are collectively responsible for the outcome [10] and individual errors in judgment have significant consequences [51]. As individual errors can incur significant cost to all members, ensuring that members have a consensus on the final sequence is crucial. However, in groups where the leader or specific individuals will be more responsible for a decision (e.g., parents' expenses during a family vacation with their young children) consensus building may not be suitable as only on a subset of the group members are subjected to the costs of members' collective opinions. Additionally, consensus building requires that members have a high willingness to discuss and negotiate to resolve potential conflicts [65]. Thus, groups with significant status disparities [57] or disagreements regarding the status of members [36] are not suitable for consensus building as members may not be willing to contribute to the discussion or be unresponsive to others' contributions. Finally, due to the time necessary to build a consensus, applying the process in urgent or emergency scenarios can be more detrimental when compared to faster decision-making methods such as majority voting [57]. Thus, in our work, we consider non-urgent CoSeq tasks in which group members are equal in responsibility and status.

Additionally, fundamental challenges can impede its successful execution. For successful consensus building, members must express their opinions regarding proposals and, then, identify disagreements that may exist between members' opinions and resolve these through discussion [5]. However, these processes can require significant time and effort in communication [10]. The laborious nature of the process may lead groups to avoid or ignore conflicts which leads to *false consensus* [18], where members hold disagreements with the final decision but failed to address them during the discussion process. Besides the effort required, various social factors may further hinder the effectiveness of the process. For instance, effective consensus building requires members to adequately consider each others' opinions [10]. However, through the anchoring effect, members may become "anchored" to their initial opinions and remain unaffected from learning about others' opinions [21, 60]. Additionally, Avery et al. [5] advised that it is essential to maintain an accessible discussion during consensus building such that all members can participate and express their opinions. Discussions, however, have been shown to easily become dominated or affected by one member or a small subset of the group [33, 41]. For example, Stettinger et al. [56] showed that the first member to express their preferences had a detrimental influence on the consequent discussion. To mitigate these challenges that surround the process, consensus building literature recommends the use of a facilitator [10]. However, as effective facilitators or leaders may not be present in all types of groups, our work instead explores the application of visual awareness to facilitate the consensus building process in CoSeq tasks.

2.3 Approaches for Supporting Group Consensus Building

Due to the merits of consensus but the difficulty in effectively building it, previous work has introduced system support to facilitate the process. Inspired by previous work which supported the tasks of individuals by visualizing the behaviors of other users [62] and supporting *social navigation* [12], a substantial amount of work has explored the use of visual awareness for group tasks [24, 26, 27, 32, 61]. For example, Goyal and Fussell [21] provided visual awareness, in the form of *sensemaking translucence*, to allow group members track their whole group's activities during collaborative analysis. This visual awareness approach has also been adopted to support consensus building by visualizing opinions in groups [2, 49, 69]. Although these systems effectively supported consensus building by increasing awareness of others' opinions and facilitating the identification of conflicts, all of them supported tasks that involve deciding on a single item. For example, Hong et al. [29] extended *dynamic queries* [1] to visualize the preferences on criteria set by members of a group to facilitate the collaborative filtering and selection of a location from a list of alternatives. Similarly, *ConsensUs* [43] facilitated the process of selecting a candidate for an engineering school by allowing the user to rate alternatives based on criteria, and to identify conflicts with their group members' ratings by visually comparing the ratings. As CoSeq involves multiple decisions, these approaches would demand significant effort from groups to iteratively specify preferences for each decision. Furthermore, as decisions in CoSeq are interrelated since they come together in one sequence, iterative decision-making may obscure these relationships. Inspired by previous work but noticing the unique challenges of CoSeq, we formulate that individually constructed sequences can be leveraged to represent members' opinions, and that facilitating the visual comparison of these can support visual awareness and consensus building in CoSeq.

2.4 Techniques for Sequence Comparison

Our work investigates supporting visual awareness in CoSeq through sequence comparison. As sequences are a subclass of graphs, comparison techniques developed in the field of graph theory could be applicable to sequence comparison. For instance, graph edit distance is a common tool used to quantitatively measure the similarity between graphs [7, 19], and adjacency matrices are frequently used to support visual comparison [3]. However, as these were designed for more general structures, they might be unable to capture and leverage the characteristics specific to sequences—such as their unidirectionality, and the importance of the relative position of nodes in the sequence. In contrast, one line of work investigated the design of visualization tools specifically for large sequence datasets [15, 44, 47, 63, 68]. For example, Monroe et al. [47] developed a system that aggregates and visualizes noisy sequence data from electronic health records. While these techniques are effective for the identification of general trends in large amounts of sequences, they cannot adequately support the exploration of detailed similarities and differences between sequences. In CoSeq, detail is necessary to allow groups to develop a comprehensive awareness of members' preferences. Most relevant to our work is *Delta* [40], which supported sequence comparison by merging them through the technique by Andrews et al. [4]. However, *Delta* only allowed for the comparison of two sequences at a time. Extending its support beyond pairwise comparison can be a challenge due to the greater difficulty of comparing three or more items [20].

In summary, findings in CSCW, HCI, and Information Visualization (InfoVis) communities show that visually externalizing group members' preferences and actions in group work, such as collaborative data analysis [27] and collaborative information-seeking [29], could lead to better outcomes. However, we identified that there is a lack of systemic and dedicated approaches aiming to understand how to specifically support CoSeq. To bridge this gap, we conduct a formative study to gain a deeper understanding of improving CoSeq through visual awareness.

3 FORMATIVE STUDY

Inspired by previous work, we postulate that considering the right strategy when designing groupware with visual awareness can improve the outcomes of CoSeq tasks. This is because an effective design can help the user to more easily compare their own preferred choices with those of their group members and allow the group to determine a sequence that reflects every necessary perspective from the various stakeholders. To educate our design for visual awareness in CoSeq, we conducted a formative study.

3.1 Method

Through our literature review, we identified four techniques that could be adapted to satisfy the requirements of CoSeq—specifically, allow for the simultaneous and detailed comparison of more than two sequences. Then, we designed prototypes based on each of these techniques and adapted them to support CoSeq (shown in Fig. 1). Below, we provide descriptions of the four prototypes:

- **Lists of nodes and edges**: this presents a list of *nodes* (i.e., selected items from a user) and a list of *edges* (i.e., pairs of adjacent nodes a user connected) with visual cues that show which nodes and edges are agreed-upon by which members of a group through the color-coding technique by Hong et al. [29] (vertical bars in the list of nodes and edges in Fig. 1 (a) show the visual cues).
- **A merged graph**: based on the technique by Andrews et al. [4], this prototype merges multiple sequences in a single graph, applying color-coding to present a group's node and edge selections (see Fig. 1 (b)).
- **A visualized edit distance**: this prototype displays group members' sequences in a row and visualizes the operations in the edit distance (i.e., add, remove, and move) between sequences next to each other in the row (see Fig. 1 (c)).
- **An adjacency matrix**: an adjacency matrix that visualizes each member's "from" choices (rows in Fig. 1 (d)) and "to" choices (columns in Fig. 1 (d)), based on Alper et al. [3]'s work that demonstrated the effectiveness of adjacency matrices for graph comparison.

With these four prototypes, we conducted a formative study to understand their strengths and weaknesses. Specifically, our primary goal was to understand how effective each prototype was at granting the user awareness of preferences within a group, as well as the user's challenges and potential needs during this process. To concretize a domain for the study, we selected travel itinerary planning among the possible CoSeq domains. This domain was chosen for this study—and the subsequent summative study–as travel itinerary planning is a commonly occurring task in real-life scenarios [30], and we wanted to ensure that we could recruit an adequate number of participants with sufficient experience, expertise and interest in the chosen CoSeq task. Additionally, previous work has shown this task to have complexity beyond being a travelling salesman problem, due to the emergence of conflicts between people's preferences [67]. In our study, we showed the four prototypes and observed how our participants used it. We recruited participants gradually until we were able to determine patterns that show general difficulties and needs. As a result, we worked with four participants who had experience in travel planning in groups (age M=22.8, two males and two females).

In this study, participants were asked to individually compare three itineraries constructed by three fictitious travelers, and construct a sequence by considering the preferences of all the group

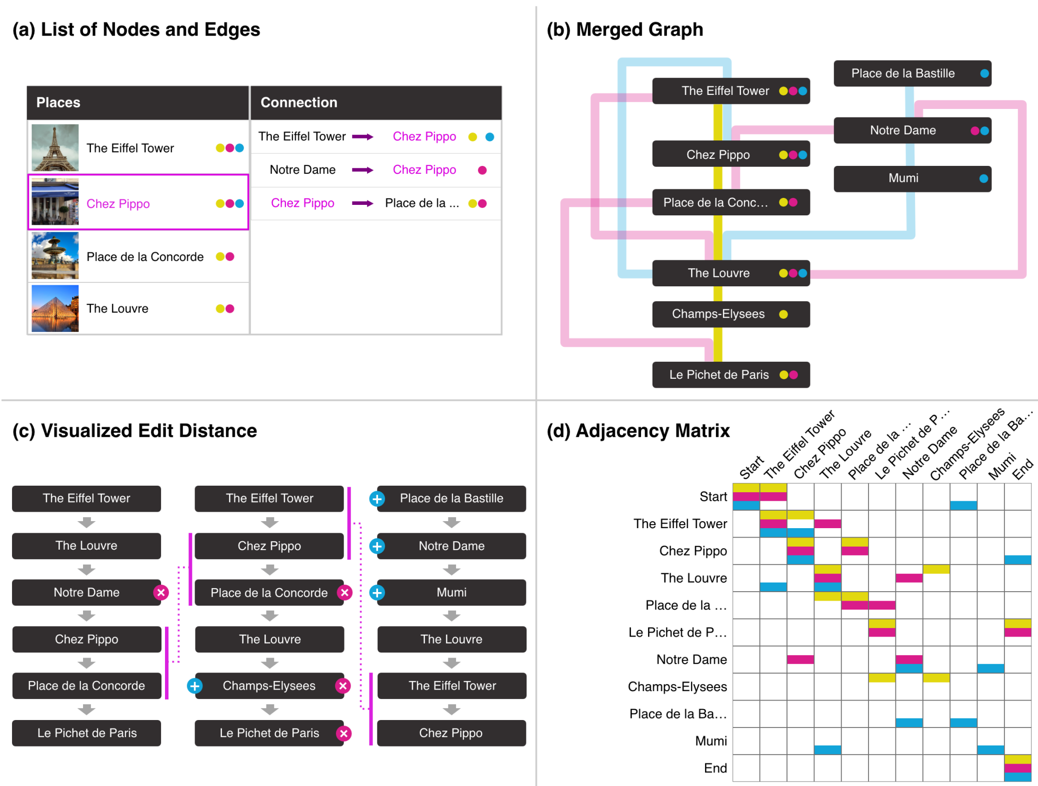

Fig. 1. The four prototypes: (a) list of nodes and edges; (b) merged graph; (c) visualized edit distance; and (d) adjacency matrix.

members. The participants were allowed to freely use any of the prototypes to evaluate and compare the given itineraries. Designed as a think-aloud study, the participants were encouraged to vocalize their thoughts and reasoning for their actions throughout the task. After the task, we conducted semi-structured interviews for an hour in which participants were asked to reflect on the strengths and weaknesses of each prototype, and comment on additional support they needed but that was not provided by any of the prototypes. Participants were also provided with a separate sheet which contained the monetary costs of all the points-of-interest (POIs) present in the three itineraries and a map which showed the location of each POI. When constructing the sequence, participants were asked to maintain the total cost of selected POIs below a specified maximum value and to ensure that the route of POIs was realistic—considering the distance between POIs and without significant winding. These factors were enforced on participants to recreate real-world CoSeq scenarios in which the weight of edges (i.e., distance between locations), ordering of nodes, and overall sequence constraints must be considered.

Participants' comments throughout the sessions were recorded and transcribed, and observations of participants' use of the four prototypes were also noted. To analyze this data, we conducted an iterative qualitative coding process. Informed by Saldaña's comprehensive manual on qualitative coding methods [53], one of the authors first applied an *initial coding* method to gain an in-depth understanding of the data and its nuances. After initial similarities and differences within the data were identified, a *pattern coding* method was applied as the initial codes showed patterns in participants' behaviors. An additional author revised the initial codes and then participated in a discussion with the initial author to perform the pattern coding.

3.2 Design Requirements

Through the analysis of participants' comments and behaviors, we identified the life span of how a user builds a consensus in CoSeq: (1) identifying *selection agreements*, (2) focusing on *ordering disagreements*, and (3) deciding on which *actions* to take to gradually reach a consensus. For each stage, the formative study findings granted insights into why and how visual awareness is leveraged in each stage. Also, we discovered that no single prototype could adequately support the entire life span. Based on these insights, we derived three high-level design requirements which could serve as a useful framework for designers who aim to facilitate consensus building in CoSeq through visual awareness (DR1, DR2, and DR3 listed below).

DR1: Display Selection Agreements at a Glance: We found that participants appreciated having visual awareness in CoSeq because it enabled them to spot what aspects in a sequence were agreed-upon by the group. Specifically, all four participants first identified nodes and edges (i.e., pairs of consecutive nodes) which were selected by multiple group members. For this purpose, they relied on the prototypes which allowed them to identify these similarities at a glance—such as the lists of nodes and edges, and the adjacency matrix. On the other hand, the prototypes which compared sequences in detail—such as lines connecting different nodes in a merged graph—were less useful. For example, P1 mentioned, *"I like [the list of nodes and edges] as it easily shows what [nodes are] popular without adding too much information."* This aspect suggests the importance of designing visual awareness in CoSeq such that the user can immediately and easily capture selection agreements between sequences. Although this resembles Shneiderman's visual information seeking mantra [54]—*"overview first, zoom and filter, then details-on-demand"*—our findings provide specific details on the type of overview needed in CoSeq tasks: an overview of selection agreements.

DR2: Present Ordering Differences on Demand to Induce Actions for Consensus: After identifying agreed-upon selections between the sequences, we found that participants used these parts as anchors to identify differences in ordering—i.e., conflicts that needed to be addressed. In particular, we found that participants characterized these differences into two types: (1) nodes or edges that appear in several members' sequences but that differ in their positions in each sequence (i.e., near the beginning, middle, or end of a sequence), and (2) agreed-upon parts in the sequences that have different nodes adjacent to them. For this purpose, participants mentioned the merit of prototypes that provided detailed information on these types of ordering differences. For example, P2 and P3 used the move operations in Fig. 1 (c)—the lines that show how the position of a node should be moved in a sequence to match another sequence—to see the distance between the positions of the same node in different sequences. In turn, this visual awareness of ordering differences allowed participants to realize what the conflicts obstructing consensus were and between which sequences they were present. This pattern of behavior indicates the design of visual awareness in CoSeq should allow users to *"zoom and filter"* on agreed-on selections and obtain *"details-on-demand"* regarding the ordering differences. Again, while these suggestions resemble those presented in Shneiderman's mantra [54], our findings present specific details to satisfy the mantra in CoSeq tasks.

DR3: Display Actions for Conflict Resolution and their Consequences: After identifying disagreements, a common pain point that participants mentioned was to determine the possible ways in which the disagreements could be resolved. Specifically, they wanted to know how specific changes to one sequence would affect the agreement between all the sequences. For example, P2 mentioned, *"I wish I could make Photoshop-like layers so I could quickly test a change on one [sequence] and see how that would change how similar the [sequences] are to each other."* This type of visual awareness was required in order for participants to easily evaluate whether a certain action they planned to take would increase or decrease the degree of agreement within a group, and how significantly. This aspect of visual awareness has not been discussed in previous work. None of the four prototypes effectively afforded as such; the prototypes present visual awareness about agreements and disagreements of multiple sequences rather than providing visual awareness on how the sequences could be changed and how such changes would affect the group's consensus. This finding implies that the design of visual awareness in CoSeq must also facilitate possible actions and explain the consequences of these actions. Fulfilling this requirement can allow a user to leverage visual awareness to clearly identify effective actions for reaching a consensus.

4 TWINE: IMPLEMENTING VISUAL AWARENESS FOR COSEQ

To fulfill these three requirements, we designed a novel visual awareness technique for CoSeq and instantiated this design by constructing Twine, a web-based research prototype designed for travel itinerary planning. Before describing our visual awareness technique, we illustrate the system to contextualize the design. Twine consists of two screens: the individual screen and the collaborative screen. In the individual screen, the user independently constructs their preferred sequence. This step was included as previous work [46, 56] has noted that it is preferable for members to form their preferences independently from their group. Next, the user and their group members proceed to the collaborative screen through which they compare sequences and discuss to reach a consensus, with the support of our visual awareness technique.

Below, we walk through a scenario which first presents the details of Twine, and then illustrates the functionalities and advantages of our visual awareness technique.

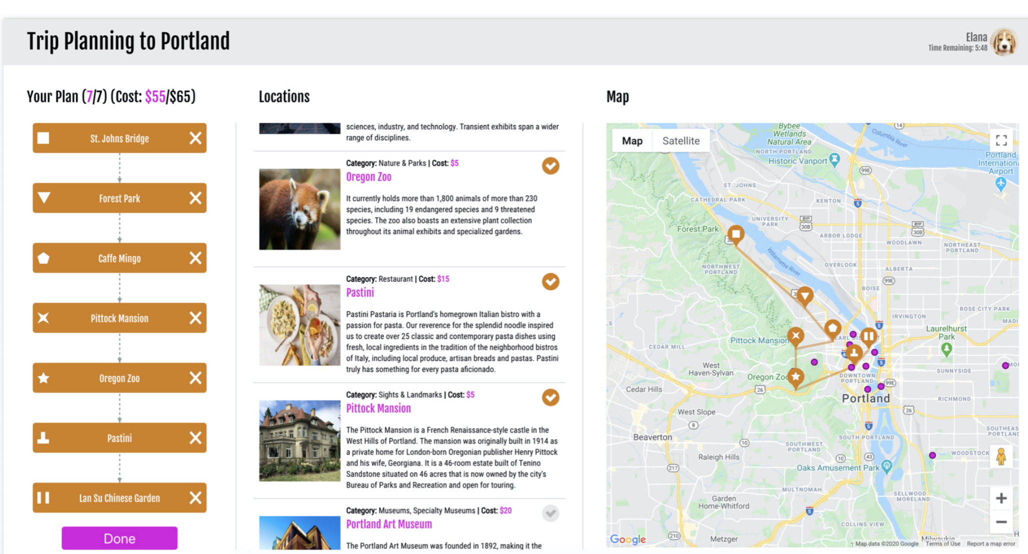

Fig. 2. The individual screen in Twine. From left to right: the sequence space, the list of POIs, and the map.

4.1 Individual Screen

Elana is planning to go on a weekend trip to Portland with her two friends, Santiago and Annie, and she wants to use Twine for this purpose. After accessing the website, she creates a session and is provided with a unique URL for her group's session which she then shares with her friends.

Once all of her friends access the URL she shared, Elana proceeds to the individual screen (Fig. 2) and her friends also proceed to their own individual screens. In this screen, Elana starts browsing through the list of points-of-interest (POIs) available in the city. As she scrolls through the list, she reads details on each POI and, if they appeal to her, she adds them to her sequence by clicking on the list entry. She can see all the POIs she has added in the sequence space, with each POI represented by a rectangular node labeled with the location's name and a unique icon.

After Elana finishes selecting all the POIs that she liked, she decides to look at the map to see if their ordering is appropriate. She notices markers in the map which are labelled with the icons representing each POI in her sequence, and lines illustrating the path of her sequence. By looking at the map, Elana realizes that "Caffe Mingo" (pentagon icon) is far from "Forest Park" (triangle icon), the location she planned to visit before the cafe. However, she also realizes that "Caffe Mingo" is close to "Lan Su Chinese Garden" (pause icon), which she planned to visit last. Therefore, in the sequence space, she drags-and-drops "Caffe Mingo" to the space after "Lan Su Chinese Garden". This changes the ordering of Elana's sequence and, as she is satisfied with the result, she submits her sequence by pressing the "Done" button. After waiting for all of her friends to also submit their sequences, Elana and her friends proceed to the collaborative screen.

4.2 Collaborative Screen

The collaborative screen (Fig. 3) is similar to the previous screen, except that Elana notices the chat feature at the bottom right, which she uses to greet her friends, Annie and Santiago (Fig. 3a). She also notices that she can now see her friends' sequences next to her own sequence in the sequence space (Fig. 3b). She recognizes which sequence belongs to each friend by the label at the top of each, and sees that each friend has been assigned a unique color. Elena can modify her own sequence and, while she cannot modify her group members' sequences, she can see modifications they make in real-time. In this screen, Elana and her friends must discuss in the chat and modify their individual sequences. The system confirms that the group has reached a consensus once all the members' individual sequences are the same. However, Elana finds it difficult to inspect her friends' sequences and understand whether they have similar or different preferences to hers. Therefore, she explores the additional support provided by the system, the visual awareness technique.

4.3 Visual Awareness Technique for CoSeq

The three design requirements identified through our formative study led to the design of our visual awareness technique, which consists of three main components: *color-coding of shared nodes*, *ordering details on hover*, and *list of missing nodes*. The color-coding of shared nodes displays colored tabs next to the user's selected nodes which indicate which of the other members have also chosen the same node. This satisfies DR1, *"provide selection agreements at a glance"*. If a node has only been selected by the user, it is colored gray to signify that the user must remove it to reach a consensus—satisfying DR3, *"display actions for conflict resolution and their consequence"*. If the user hovers over a node in their own sequence, the ordering details on hover highlights that node in the other members' sequences, if they also selected it, to facilitate comparison of relative positions. Also, it displays the actions the user can take to match members' sequences with respect to nodes adjacent to the hovered over node. This satisfies DR2, *"present ordering differences on demand to induce actions for consensus"*, and DR3. Finally, the list of missing nodes displays nodes that the

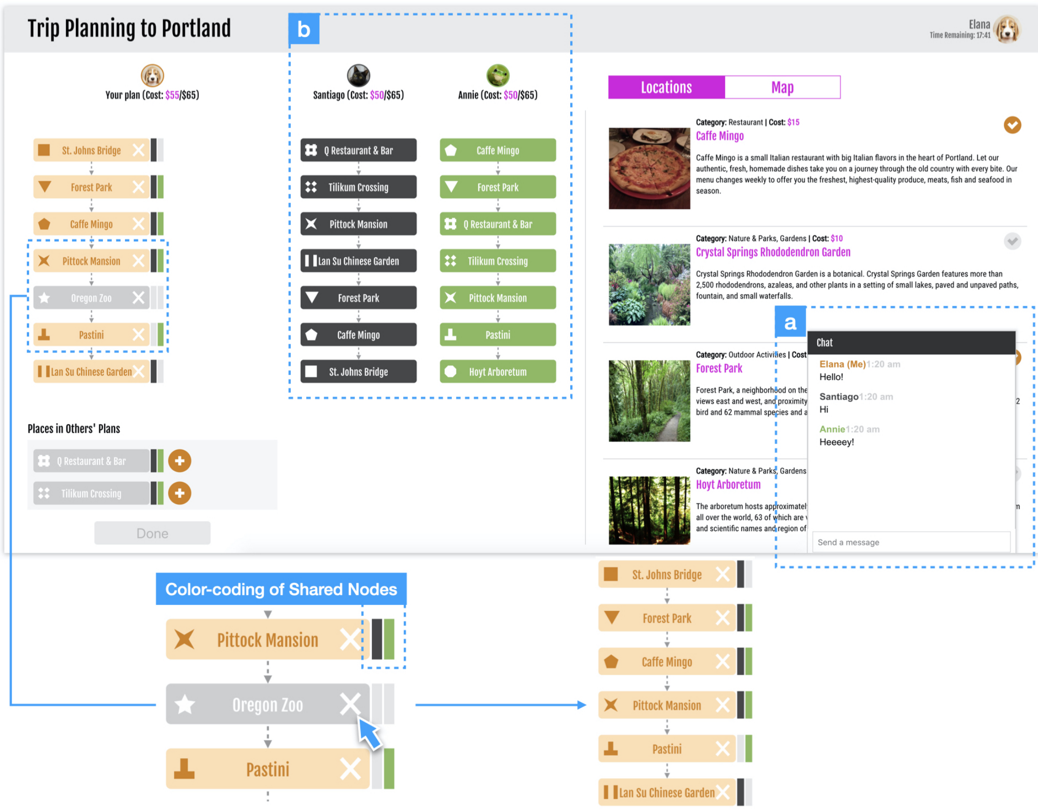

Fig. 3. The collaborative screen of Twine presents a chat window (a), the user's sequence, her members' sequences (b), and the visual awareness technique. The figure shows the user using the *color-coding of shared nodes* to identify that only she selected the "Oregon Zoo" node, and then removing that node from her sequence.

user did not select but their group members have, and allows the user to easily see which members have selected it—satisfying DR1—and add these nodes to their own sequence—satisfying DR3. We illustrate each component in detail through our user scenario and figures below.

Color-coding of Shared Nodes (Fig. 3): If Elana wants to understand which of her friends also selected a node that she selected, she can examine the colored stubs to the right of that node in her sequence (DR1). For example, Elana looks at the node for "Pastini" in her sequence, and sees that one stub is colored green, the color that represents her friend Annie, which shows that Annie also selected "Pastini". However, the other stub is colored gray which shows that her friend Santiago did not include it in his sequence. Additionally, Elana notices that her node for "Oregon Zoo" is fully encoded in gray. This allows her to easily recognize that she was the only one that selected the zoo (DR2) and that, by removing it, she can come closer to an agreement with her friends (DR3). Knowing this, she removes the zoo from her sequence by clicking on the 'X' mark on that node.

Ordering Details on Hover (Fig. 4): Elana notices that both of her friends also selected "Forest Park" in their sequences. Her group must now agree on when to visit this park. Therefore, to compare her ordering of the park to how her friends ordered it in their sequences, Elena hovers over the park's node. This highlights that node in her friends' sequences which allows Elana to

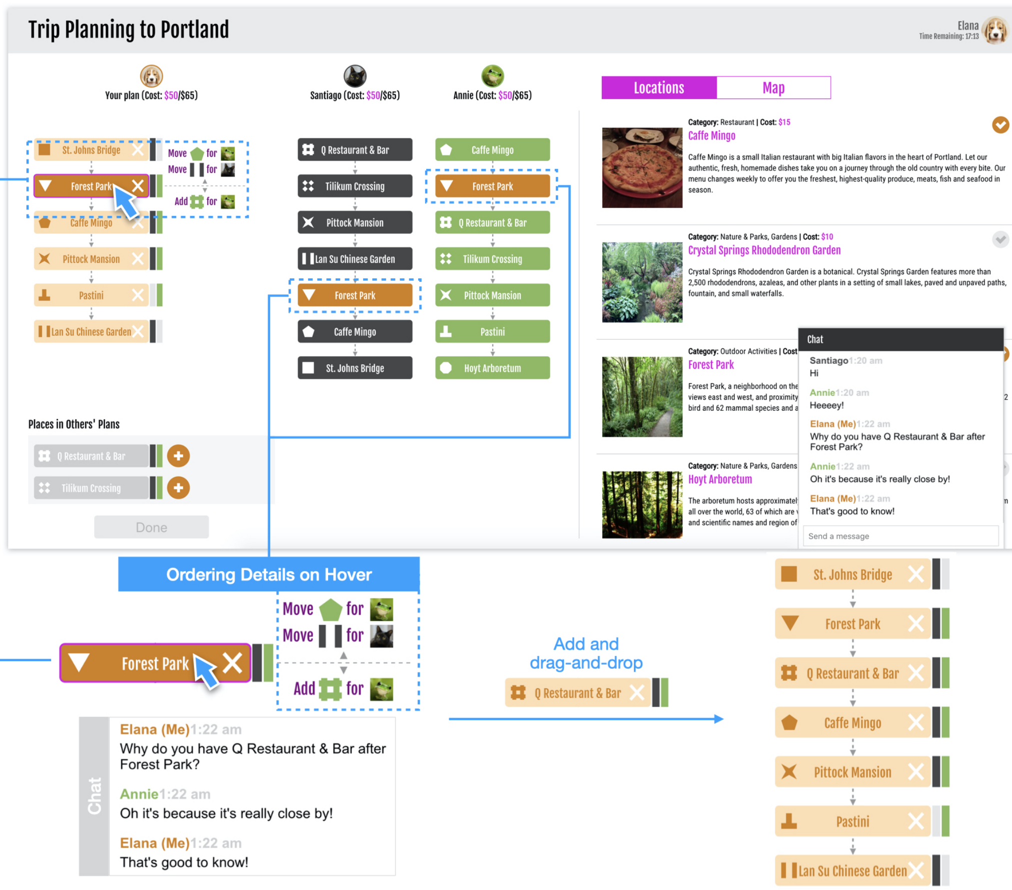

Fig. 4. With the *ordering details on hover*, the user can compare the relative positions of a node, through the highlighting, and the adjacent nodes, through the details displayed next to the node. With this information, the user discusses nodes adjacent to the "Forest Park" node with her group, and decides to add and position the "Q Restaurant & Bar" node in her sequence.

easily see that Annie positioned it close to the beginning in her sequence like Elena did (DR1), but Santiago positioned the park closer to the end of his sequence. Satisfied by seeing that she at least agrees with Annie, she now decides to compare the POIs she planned to visit before and after the park with those her friends planned to visit before and after. In other words, she wants to compare the nodes adjacent to the park's node. To do this, Elena hovers over the park's node to bring out details regarding adjacent nodes. Elana sees that, while she included "Caffe Mingo" (pentagon icon) after the park, her friend Annie had "Q Restaurant & Bar" (hash icon) after the park (DR2). In the chat, Elana asks Annie why she has "Q Restaurant & Bar" after the park, and Annie explains that it is closer to the park. With the details displayed, Elana sees that, to follow Annie's recommendation, she must add the restaurant to her sequence, as she had not selected it, and move it to the position after the park (DR3).

List of Missing Nodes (Fig. 5): With the *color-coding of shared nodes*, Elana can determine which of her friends also selected nodes that she selected. However, using only that component, she is unable to easily recognize nodes her friends selected but that she did not select. For this

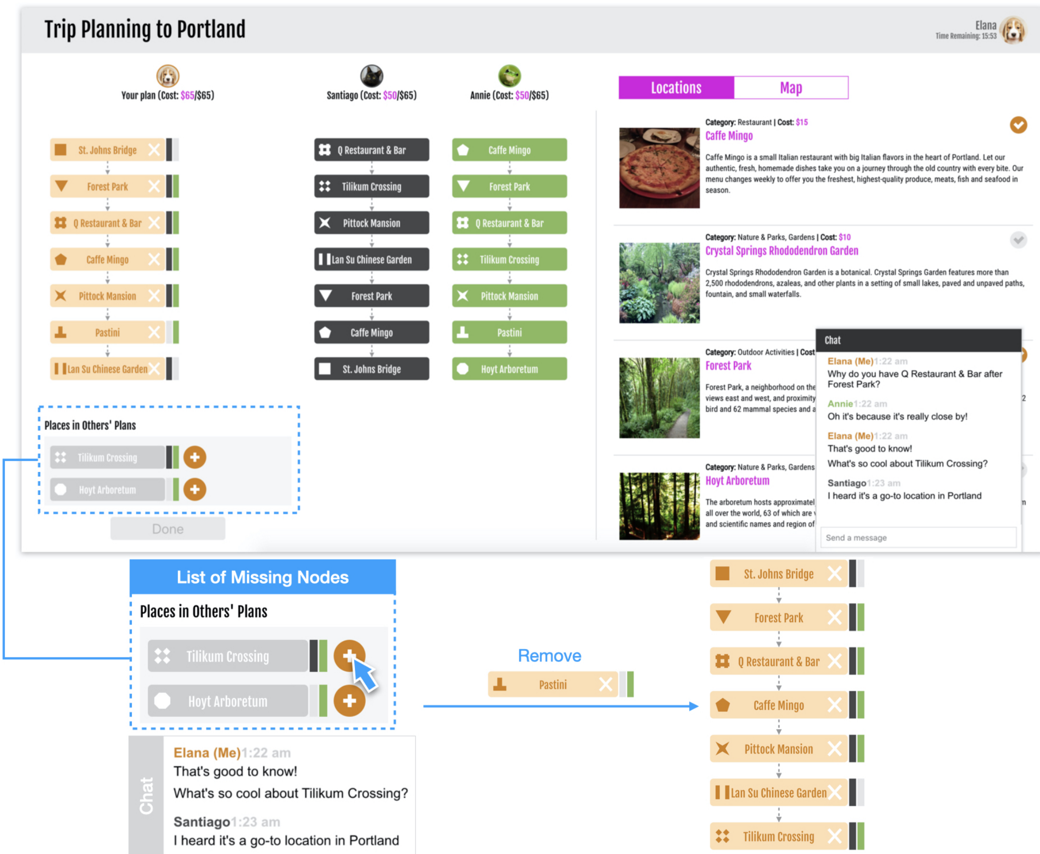

Fig. 5. The *list of missing nodes* is displayed below the user's sequence. With the list, the user notices that her two group members selected the "Tilkum Crossing" node but she did not. After a discussion about this node, she removes the "Pastini" node from her sequence to accommodate space, and adds the "Tilkum Crossing" node to her sequence by clicking the plus button next to that node in the list.

purpose, Elana uses the *list of missing nodes* which is shown below her sequence in the screen. By scrolling through the list, Elana can easily see which locations her friends have included in their sequences (DR1) and can also see how popular each one is with the colored stubs—these function in the same way as those used for the *color-coding of shared nodes*. With the list, Elana notices that both of her friends have included "Tilikum Crossing" in their sequences and that, by adding it, her preferences would become closer to her friends' (DR3). After discussing it with her friends, she agrees that visiting "Tilikum Crossing" will be enjoyable and therefore adds it to her own sequence by clicking on the plus button in the list.

4.4 Implementation

We implemented Twine's interface with Javascript, ReactJS, and CSS. The backend was implemented through a Node.js server and MongoDB database. To allow for chat messages and sequence changes to be communicated in real-time between group members, we used the Socket.io library.

5 SUMMATIVE STUDY

With our system Twine, we conducted a controlled lab study to investigate the effect of our visual awareness technique on the efficiency and effectiveness of the consensus building process of groups during a CoSeq task. As previous work has demonstrated that visual awareness can facilitate the process of establishing a common ground in other group tasks [29], we hypothesized that visual awareness can reduce the time and effort in reaching a consensus in CoSeq—i.e., increase efficiency. Also, previous work showed that visual awareness can encourage group members to be more considerate of others' preferences [49]. Thus, we believe that providing visual awareness in CoSeq can also encourage groups to incorporate the preference of more members and increase overall satisfaction regarding the process and outcome—i.e., increase effectiveness. Therefore, our hypotheses are:

- H1: Providing the visual awareness technique will reduce the time and effort required to build a consensus in CoSeq and increase participants' perceived efficiency regarding the overall process.
- H2: Providing the visual awareness technique will increase participants' perceived effectiveness regarding the consensus building process and their individual satisfaction towards the final sequence constructed.

5.1 Method

We present details regarding the setting and procedure of our study.

5.1.1 Participants.

We recruited 15 groups of three friends (age M=22.0, SD=1.81, 37 males and 8 females). The participants applied in groups of their choice. All 45 participants were unique—none of them participated in the study more than once. Each participant was compensated KRW 20,000 (~$17.00) for a 90-minute task. We controlled for possible factors related to group dynamics which could have had a significant influence on the consensus building processes of groups: the number of members in each group was constrained to three to control for group size; only groups of friends were recruited to control for hierarchy among members and relationship strength; and only students from a local university were allowed to participate to control for homogeneity within groups. Recruitment was carried out through online forums. Additionally, we chose a synchronous and dispersed setting for all the groups: group members participated in the task simultaneously, but were physically separated to prevent verbal communication. This setting was chosen to prevent delays in communication due to asynchronicity and to limit the avenues of communication to chatting through our system.

5.1.2 Apparatus.

We conducted a within-subjects experiment by manipulating visual support. A within-subjects design was taken as it is commonly applied in group studies [16, 29, 42]. Also, subjecting each group to all conditions reduces possible noise on the observed effect attributed to factors related to individual group members (e.g., preferences) and the relationships between group members. Two versions of Twine were compared (shown in Fig. 6): (a) Control condition: a baseline version which only juxtaposed the members' sequences in the sequence space of the collaborative screen, and (b) VA condition: a version with the visual awareness technique embedded in the sequence space. The two conditions differed only in the sequence space of the collaborative screen. The individual screens and the other components in the collaborative screens (i.e., list of POIs, map, and chat) were the same. To allow for a fair comparison between the two conditions, we chose two cities in the US, namely Portland (Oregon) and Denver (Colorado), which are similar in terms of area, population, and popularity (in terms of the total number of reviews on the travel website TripAdvisor). All participants reported to have never visited or lived in either city. For each city, we prepared equal-sized POI datasets that contained information for a total of 20 POIs: 14 attractions, and 6 restaurants or cafes. POIs were chosen based on their high number of reviews on TripAdvisor. For each POI, five types of information were obtained from various online sources: name, categories, a short description, coordinates, and approximate cost. Through our unstructured interview prior to the study, these five were identified as the most important factors users consider when browsing through and deciding among POIs. The information of each POI was obtained through the TripAdvisor entry for the POI, tourists' text reviews (especially to obtain the approximate costs of restaurants or cafes), or the POI's official website.

5.1.3 Study Procedure.

The study took place in a computer lab. Each participant was assigned a computer and participants in the same group were kept at a distance from each other. After reading and signing the informed consent form, each group was asked to perform the same task twice: to construct and reach a consensus on a travel itinerary, once with the Control condition and once with the VA condition. Before each trial, they were provided with an overview of the experimental procedure and a brief walkthrough of the version of the system to be used for that trial of the task. We verified that participants thoroughly read the walkthrough documents to ensure that they were aware of the system features included or excluded in each condition. After each trial, participants were asked to answer a short survey. To mitigate the influence of ordering effects, the order of the conditions was counterbalanced by randomly assigning the groups into two sets—one of seven groups and one of eight—and then randomly assigning the sets to start with the Control or VA conditions. This resulted in eight groups starting with the Control condition, and seven groups starting with the VA. Regardless of their condition ordering, all groups considered POIs in the Portland dataset in their first trial and then those in the Denver dataset in their second trial. The aim of using two datasets

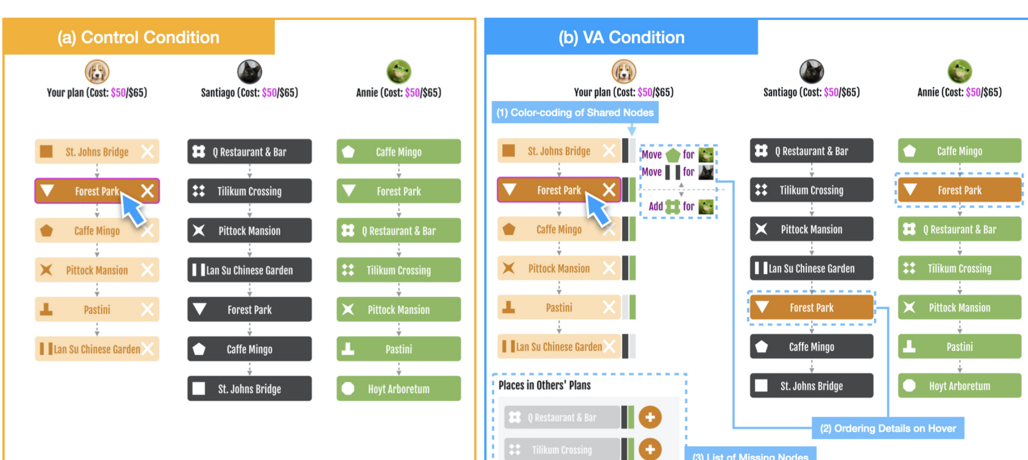

Fig. 6. The two versions of Twine used as conditions in the study differed in the sequence space. The Control condition (a) only showed the members' sequences in a row, without the additional interactions supported by our visual awareness technique. The VA condition (b) incorporated the visual awareness technique: color-coding of shared nodes (1), ordering details on hover (2), and list of missing nodes (3).

and fixing their order was to remove learning effects on the datasets between the two trials and to separate the datasets from the two conditions.

The task consisted of two phases. In the first phase, each member used the individual screen to construct their preferred sequence, which had to include five to seven POIs. Once all members submitted their sequences or the assigned time of 10 minutes for the phase was completed, the group progressed to the second phase. In the second phase, members used the collaborative screen to discuss, and compare and modify their individual sequences to reach a consensus. The second phase could be completed once all members shared the same sequence, which would make a "Done" button clickable. A randomly chosen member in each group would be the only one to be able to click this button to complete the phase. This member's only responsibility was to check one last time, before clicking, that all group members were satisfied with the currently shared sequence. Participant groups had 20 minutes to complete the second phase—a time decided through a pilot study. All groups reached a consensus within the assigned time and, on average, did this within approximately 14 minutes showing that the time provided was sufficient.

During the two phases of the task, participants were asked to imagine that they would actually travel to the cities in question, and consider their own sincere preferences in relation to the available POIs. Additionally, they were asked to construct routes of POIs in their itineraries which they would be realistically willing to follow, and a maximum budget was enforced on the total cost of POIs included in a sequence. Similar to the formative study, participants were provided with these instructions and constraints to recreate factors in real-world CoSeq scenarios. By looking at participants' messages, we verified that groups actively discussed about these factors—for example, messages like *"isn't the museum closer"* or *"let's not go to the mountain, it's too far away"* were common.

5.1.4 Measures.

We used mixed methods, integrating both qualitative and quantitative data. For qualitative data, we collected participants' responses to the following open-ended survey questions: *"Which of the two versions of the interface better supported your group to reach a consensus and why?"*, and *"What changes could be made to the interface to help you perform this task more efficiently or effectively?"* For quantitative data, we considered both survey responses and system logs. Additionally, a discourse analysis of the chat logs was conducted to gain a detailed understanding of the participant groups' discussions in each condition. Two of the authors independently open coded all the chat logs (7,251 messages) into potential message classes. Then, the two authors met to merge and narrow down classes to synthesize a final classification scheme composed of nine classes (see Table 1). Using this scheme, they jointly coded two chat logs (408 messages, 5.63% of the total) to refine and reach a mutual understanding on the classes. Finally, they independently coded the remaining chat logs (Krippendorff's alpha=0.752) and then discussed to resolve conflicting codes. Our discourse analysis was based on the method used by Lee et al. [42] to tag the intents of messages by moderators in chat-based conversations. As this allowed them to observe differences in the discussion behaviors of moderators, we saw this approach to be suitable to observe condition-based differences in the participant groups' discussions. As discussion is a crucial part of the consensus building process, understanding how the discussions differ between conditions can provide detailed insights regarding the effect of the visual awareness technique on consensus building in CoSeq.

We evaluated the efficiency and effectiveness of the consensus building process through both attitudinal and behavioral measures. Attitudinal measures refer to those which evaluate the participants' subjective opinions and perceptions, while behavioral refer to those which involve observing participants' actions during the process.

Class	Explanation	Example	Percentage
Express Opinion	Stating their subjective preferences in general, or regarding a specific instance (i.e. node or ordering).	• *"I like looking at animals."* • *"Oh the zoo looks pretty! I like the zoo."* • *"Eating right after going to the zoo sounds gross…"*	50.7%
Ask for Opinions	Asking a specific group member or the group in general for their preferences.	• *"Do you guys like museums?"* • *"What are your opinions on eating after the zoo?"*	13.4%
Propose a Node Selection Decision	Proposing a specific node (i.e. location) for the group to include or exclude from their sequences.	• *"Let's go to the Transport Museum."* • *"We have to go to the zoo."*	13.4%
Propose a Node Ordering Decision	Proposing a specific ordering for the group members to arrange certain nodes in their sequences to.	• *"Let's change the Transport museum so it goes after the zoo."* • *"What if we go to the restaurant for dinner instead of lunch?"*	8.54%
Manage the Discussion	Messages used to manage the groups' discussion (e.g. summarizing, clarifying, calling for focus).	• *"So we decided on zoo -> museum -> lunch."* • *"Enough chit-chatter, we only have 5 minutes left."*	7.56%
Compare All the Sequences	Message comparing the sequences of all the group members.	• *"One location all of us added is the zoo."* • *"We all have the bridge last."*	2.75%
Compare a Pair of Sequences	Message comparing the sequences of two group members.	• *"It looks like [name] and [name] both added the museum."* • *"You and I both have the museum after lunch."*	1.07%
Ask or Demand a Change	Asking or commanding another group member(s) to make changes to their sequences.	• *"Hurry up and add the zoo."* • *"Could you please add the zoo?"*	1.49%
Express Willingness to Change	Mentioning willingness to change or modify their sequence.	• *"I can take out the zoo if you guys don't want it."* • *"I can add the zoo by taking out the museum."*	0.980%

Table 1. The classification scheme for messages related to consensus building in CoSeq. The scheme was used to code all the study chat logs which resulted in 2,142 codes. For each class, the table includes a percentage that shows the proportion, out of the total count of codes, for which codes from that class accounted for.

Attitudinal Efficiency: Two measures, based on post-survey questions, were used to measure participants' perceptions regarding the efficiency of their groups' consensus building processes. To measure participants' perception on the efficiency of the process, a 7-point Likert scale question in the survey asked, *"The tool helped me perform the task of planning a shared itinerary that was satisfactory to everyone with a small amount of effort."* We refer to this measure as "perceived efficiency" hereinafter. The survey also contained five questions on NASA-TLX [25], a questionnaire that measures perceived workload through six questions related to mental demand, physical demand, temporal demand, performance, effort, and frustration. We excluded the question on physical demand in our survey. By inverting the scale for the performance question and averaging the responses for each question, one measure of cognitive load could be calculated. Additionally, participants' responses to the survey's open-ended question were analyzed to extract insights on the effect that the visual awareness technique had on the efficiency of their consensus building processes.

Behavioral Efficiency: From the system logs of each group and trial, we collected (1) the total time taken to complete the second phase (i.e., the collaborative phase), (2) the total number of chat messages sent as a group, and (3) the count of codes for five message classes. Due to the substantially different proportions of each message class in the chat logs, we only considered those that were relatively more common for the analysis. The following are the five that were considered: “Express Opinion”, “Ask for Opinions”, “Propose a Node Selection Decision”, “Propose a Node Ordering Decision” and “Manage the Discussion”. The degree of similarity or difference between the group members’ initial sequences can differ between trials, and this can affect the time and effort needed for the group to reach a consensus. If the sequences were substantially different, there would be more conflicts on which the group needs to expend time and effort. The inverse would also be true—similar initial sequences would mean less conflicts, time and effort. Thus, for each trial, we normalized the efficiency behavioral measures by dividing measured values by the total initial edit distance (TIED). TIED is the sum of the edit distances between each possible pair of the group members’ initial sequences. A larger TIED implies that the initial sequences are largely different to each other, and a smaller TIED implies the sequences are similar to each other. To keep consistency with the possible actions allowed in the system, the edit distance calculation only considered add, remove, and move operations—disregarding substitutions.

Attitudinal Effectiveness: Similar to attitudinal efficiency, attitudinal effectiveness was measured through post-survey questions. To measure the participants’ perception on the effectiveness of the process, participants were asked a 7-point Likert scale question: *“The tool was effective in helping me perform the task of planning a shared itinerary that was satisfactory to everyone.”* We refer to this measure as “perceived effectiveness” or “perceived group satisfaction” hereinafter. To measure participants’ individual satisfaction towards the outcome, each participant was shown the sequence their group reached a consensus on and were asked to rate, on a 7-point Likert scale, their personal satisfaction towards it. Participants’ responses to the survey’s open-ended question were also analyzed with consideration to effectiveness.

Behavioral Effectiveness: Due to the subjectivity of exploratory tasks like CoSeq, it is challenging to quantify the effectiveness of groups’ behaviors or to specify “ground truths” to evaluate their outcomes against and measure success. However, understanding how varied group members’ initial sequences were with respect to the group’s final sequence can provide insight into how effective the consensus building process was at incorporating the preferences of all group members. For example, if the final sequence is similar to one member’s initial sequence but drastically different to the other members’ initial sequences, this could imply that only the first member’s preferences were considered thoroughly. To analyze this variation, we calculated the Gini coefficient for the edit distances between group members’ initial sequences and their group’s final sequence. A Gini coefficient closer to 0 indicates that each of the group members’ initial sequences were equally different from the final sequence, and a value closer to 1 would mean that there was a large inequality between how much each member had to change their initial sequences to reach the final sequence.

5.2 Results

For the statistical analysis of measures, we conducted either a paired t-test (if the data was parametric) or a Wilcoxon signed-rank test (if the data was non-parametric). These tests were conducted as they allow for evaluating the significance of differences between paired samples. To determine whether the data was parametric or not, we conducted normality tests through a Shapiro-Wilk test. We found a set of meaningful significant differences between the two conditions. A summary of all our meaningful results is provided in Table 2.

		Measure	Control	VA	
Efficiency	Attitudinal	Perceived Efficiency	5.21 (SD=1.05)	6.21 (SD=0.38)	**
		Aggregated NASA-TLX	3.27 (SD=1.08)	2.88 (SD=0.57)	*
	Behavioral	Time Taken (min)	0.65 (SD=0.20)	0.54 (SD=0.20)	
		Number of Messages	10.89 (SD=5.94)	8.72 (SD=5.93)	
Effectiveness	Attitudinal	Perceived Effectiveness	5.21 (SD=1.02)	6.07 (SD=0.58)	**
		Sequence Ratings	6.05 (SD=0.70)	6.37 (SD=0.38)	
	Behavioral	Gini Coefficient	0.15 (SD=0.10)	0.17 (SD=0.10)	
Number of Participants that Preferred the Condition			1	38	

* : p<.05, ** : p<.001

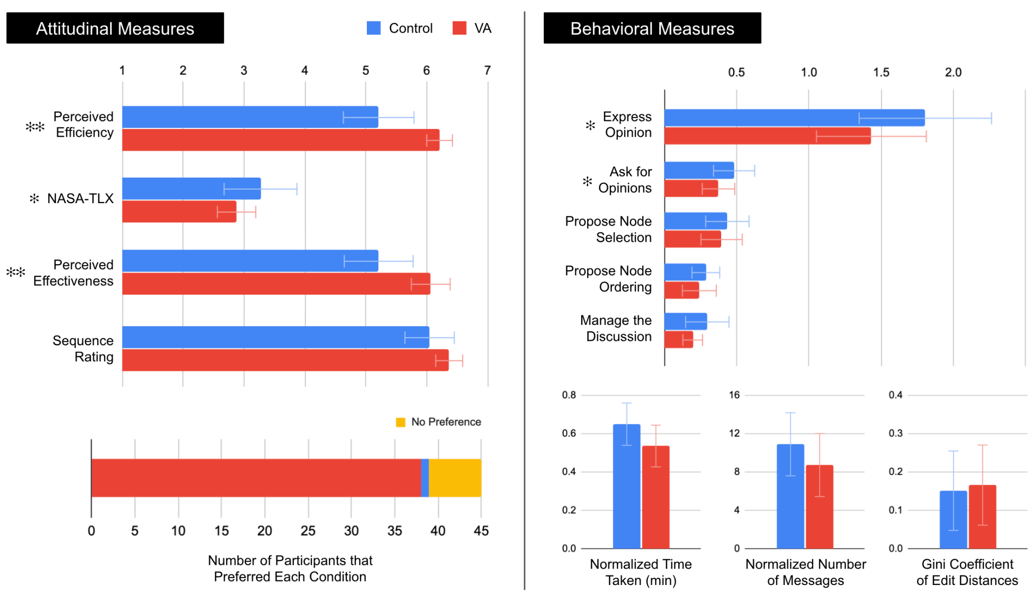

Table 2. A table (top) and plots (bottom) that summarize the results (means and standard deviations) across the two conditions and for the measures on attitudinal efficiency, behavioral efficiency, attitudinal effectiveness, and behavioral effectiveness. For the time taken to reach a consensus and the total number of messages sent, the table and plots show the normalized values. Error bars in the graphs indicate 95% confidence intervals.

5.2.1 *Efficiency.*

Attitudinal measures indicate that visual awareness can increase participants' perceived efficiency and reduce their cognitive workload. A Wilcoxon signed-rank test showed that participants perceived the consensus building process to be more efficient in the VA condition (M=6.21) compared to in the Control (M=5.21, z=28.0, p<.000, d=1.27) (Table 2, row 2). This signals that participants perceived that they could reach a consensus with less effort and time when equipped with visual awareness. Aside from perceived efficiency, participants also perceived their cognitive workload to be lower when using the visual awareness technique. This is evidenced by the results of a Wilcoxon signed-rank test of participants' aggregated NASA-TLX scores. These scores showed that participants ranked their cognitive workload to be significantly lower in the VA condition (M=2.88) than in the Control (M=3.27, z=327, p<.05, d=0.45) (Table 2, row 3). This can be due to the visual awareness allowing the user to externalize their cognition—a benefit that was noted by

Message Class	Control	VA	p-value
Express Opinion	1.81 (SD=0.83)	1.43 (SD=0.67)	**0.03**
Ask for Opinions	0.48 (SD=0.26)	0.37 (SD=0.20)	**0.05**
Propose a Node Selection Decision	0.44 (SD=0.27)	0.40 (SD=0.26)	0.97
Propose a Node Ordering Decision	0.29 (SD=0.17)	0.24 (SD=0.21)	0.96
Manage the Discussion	0.30 (SD=0.27)	0.20 (SD=0.12)	0.15

Table 3. Summary of results for the normalized counts of the five most frequent message classes from our discourse analysis. Significance tests for each message class were performed through paired t-tests. Only the classes "Express Opinions" (p=.025) and "Ask for Opinions" (p=.046) showed significant decreases from the VA condition to the Control.

G2P2: *"There was no need to remember in my head which [POI] we all wanted to go to and which [POI] only I wanted to go to."*

Facilitating Identification Agreements and Disagreements: The increase in perceived efficiency can be attributed to visual awareness allowing participants to identify agreements and disagreements between their groups' preferences with more ease. Participants expressed that they could identify agreements and disagreements between members' preferences with more ease in the VA condition. For example, G7P2 mentioned that the visual support allowed them to reduce their time needed to identify agreements: *"It was easy to see at a glance all the [POIs] that I shared with my group members and this helped me save time in identifying agreement in opinions."* G2P3 explained that, in the VA condition, their effort to notice and resolve disagreements was reduced: *"[The VA condition] allowed me to see disagreements in opinion within my group as a whole and to easily correct these."*

Supporting Expression and Recognition of Preferences: Our discourse analysis results indicate that the VA condition reduced groups' efforts in discussing preferences. The discourse analysis showed that participants mentioned their own preferences with approximately 21% less frequency in the VA condition than in the Control (p=.025, d=0.47) (Table 3, row 2). They also inquired about their group members' preferences with approximately 22% less frequency in the VA condition than in the Control (p=.046, d=0.50) (Table 3, row 3). Several participants mentioned that the visual support allowed them to easily recognize their group members' preferences regarding node selections: *"With the feature, it was possible to see what other members wanted or did not want immediately. (G5P1)"* Additionally, a couple of participants noted that preferences regarding node orderings were also more easily noticed in the VA condition: *"It was easy to see the position of a [node] in my friend's itinerary by hovering over that [node] in my itinerary. (G3P3)"*

Drawbacks of Visual Awareness: However, while our analysis of time taken and the total number of messages showed reduced effort, the differences were not significant. A paired t-test analysis of the normalized time taken to reach a consensus was lower in the VA condition (M=.535) when compared to the Control (M=.650), but only with marginal significance (p=.06) (Table 2, row 4). Similar results were shown for the normalized number of messages between the VA (M=8.72) and Control conditions (M=10.9, p=.06) (Table 2, row 5). The lack of significant differences in overall time taken or number of messages—despite significant decreases in cognitive workload and messages discussing preferences—imply that visual awareness can decrease effort related to only some portions of the consensus building process in CoSeq tasks. For instance, groups dedicated relatively the same amount of effort in proposing alternatives (Table 3, rows 4 and 5) and managing the discussion (Table 3, row 6). Decreasing the time and effort related to these discussion aspects could potentially reduce the overall time and effort. On the other hand, it is possible that having visual awareness available introduced additional time and effort to the process. For example, some participants brought up usability issues regarding the visual awareness technique. G5P2 expressed that navigating the VA condition could be challenging due to the amount of information contained in it: *"Crowding [several features] into a small screen made it look somewhat chaotic."* On the other hand, G4P2 felt difficulty in learning how to use certain components: *"It was the first time using the 'add and move' feature which made it difficult to become familiar with it and properly use it within a short time."*

5.2.2 Effectiveness.
Similar to our results for efficiency, attitudinal measures show that visual awareness can increase the perceived effectiveness of the consensus building process in CoSeq. A Wilcoxon signed-rank test showed that participants perceived the consensus building process in their CoSeq task to be more effective in the VA condition (M=6.07) than in the Control condition (M=5.21, z=20.0, p<.000, d=1.03) (Table 2, row 6). Also, a paired t-test showed that participants' satisfaction ratings for their final sequences were higher in the VA condition (M=6.37), but the differences with the Control condition (M=6.04) were not significant (p=.09) (Table 2, row 7). These results show that, with visual awareness, participants perceived the consensus building process to be more effective at increasing satisfaction at the group level, but that the technique had no significant effect in their individual satisfaction. This disparity in the effect of visual awareness on satisfaction at individual and group levels could be attributed to the fact that participants' individual satisfaction ratings were relatively high (ratings higher than 6), irrespective of the condition. However, this could also suggest that a trade-off may exist between increasing the overall satisfaction of a group and increasing the satisfaction of individual members.

Members Modified Individual Sequences by Similar Degrees: Across all groups and trials, the average Gini coefficient for the edit distance between group members' initial sequences and their groups' final sequences was 0.158. As this value is relatively close to zero, this implies that there were no large variations in how much each member of a group had to modify their initial sequences to reach their group's final or agreed on sequence. Additionally, a Wilcoxon signed-rank test showed that the Gini coefficient values did not differ significantly between when participants had the VA condition (M=0.17) or the Control condition (M=0.15, z=52, p=.49) (Table 2, row 8). This result implies that members in each group made a relatively equal amount of changes to reach a consensus, regardless of whether they had visual awareness or not.

Influencing the Discussion Process: A couple of participants described how their groups' discussion processes differed in the VA condition. For example, G10P3 described that, with the visual awareness technique, their group reached a consensus by first establishing the nodes that were selected by all members, and then discussing each node selected by at least one group member: *"We tried to include all the locations that all our members selected. Then, we discussed to decide on whether to include those locations that were only selected by one or two members."* G12P1 noted a similar process: *"We fixed those POIs that were shared by the whole group and then discussed the other POIs that each member wanted."* As explained by the participants, this type of process allowed for all the selection decisions of each member to be discussed by the group.

Encouraging Preference Adjustments: An additional effect of the VA condition that was noted by several participants was that, as it visualized others' preferences, it encouraged them to adjust their opinions. For example, G11P1 noted that, after noticing their members' preferences with the visual support, they addressed these preferences in the discussion to readjust their own opinions: *"By easily knowing who wanted to go to which location, we could ask that person about their opinion and tune our own opinions accordingly."* As the visual awareness technique in the VA condition highlighted major differences between the user's sequence and their group members' sequences, G8P2 noted that this encouraged them to make modifications: *"I quickly reduced my differences* *in opinion by seeing [nodes] that I selected but none of my group members selected colored in grey."* Similarly, G13P3 explained that the *list of missing nodes* allowed them to notice nodes that their other members shared and that this led them to adjust their individual sequence: *"Having a separate list for locations that I didn't select but my members did allowed me to conveniently adjust my opinions to match my members' opinions."*

Increasing Efficiency can Increase Effectiveness: There is some evidence that, by increasing the efficiency of the process, the VA condition granted certain groups the freedom to assess their decisions more carefully. G7P2 mentioned that, while their group approached the time limit in the Control condition, their group had time left after reaching a consensus in the VA condition which allowed them to reassess their final sequence and consider more alternatives: *"With the [VA condition], we were able to reach an agreement in less time which allowed us to have the leisure to look at the plan we created once more and to consider other possible options."* Additionally, G6P1 noted that, by simplifying certain aspects of the task, the VA condition allowed them to attentively consider their satisfaction towards the ordering of nodes as they decided on their selection: *"[With the VA condition], I could look more at the map. As a result, I could see, while making the plan, whether it was following the optimal path in the map. In the [Control condition], I only realized that the path was jumbled up once we had completed the plan."*

Need for Additional Information: Lastly, several participants expressed a need for information beyond what was provided by the system. For example, G1P2 mentioned a need to see other travelers' opinions on POIs: *"It would be better if reviews from other people were included for [each POI], as it was difficult to decide just by reading the descriptions."* Additionally, several participants demanded details on transportation and travel time between POIs: *"It was a bit uncomfortable since there was no information on the amount of time needed to walk between POIs or if public transportation is available. (G13P3)"* A couple of participants requested that the system link to external resources as they felt that the information provided by the system was insufficient: *"It would be good if clicking on a POI would open a website with information about that location. (G11P2)"*

6 DISCUSSION

Despite being a subclass of collaborative planning tasks, CoSeq is complex and introduces its own unique challenges. Our system Twine introduces a novel and customized solution to this type of task by building upon social theories, prior systems, and an empirical investigation. Below, we interpret the observed effects that our system had on the efficiency and effectiveness of CoSeq tasks.

6.1 Efficiency

In terms of efficiency, our visual awareness technique appeared to reduce the effort dedicated to discussing preferences and, thereby, improved perceived efficiency. However, we observed no significant decrease in the time and effort needed to reach a consensus. We contemplate that this mixed effect is possibly due to the lack of support for identifying alternatives and the focus required to use the technique.

Overall, participants perceived the process to be more efficient and their cognitive workload to be reduced with visual awareness. By showing the user their members' preferences in relation to their own, visual awareness could allow the user to easily perceive their members' preferences on node selection and ordering. As members' preferences were more salient, group members may have felt less need to explicitly inquire about each others' preferences or mention their own during the discussion—a result reflected by our discourse analysis. The process of revealing opinions or preferences is a key step in reaching a consensus [10] and, indeed, explicit discussion of preferences accounted for most of groups' discussions—as evidenced by the message class statistics. Therefore, this demonstrates that visual awareness can increase the perceived efficiency of the consensus building process in CoSeq tasks by decreasing the effort expended in a major step of the process.

Despite improvements in perception, our results showed no significant decreases in the overall time and effort (i.e., number of messages). As mentioned by a couple of participants, using the visual awareness technique incurred additional individual effort which could have resulted in mixed-focus [23]—focus is divided between an individual task (navigating the visual awareness support) and a group task (the consensus building discussion). As a result, discussions could have been delayed as members awaited others' responses. Additionally, our technique is limited as it supports the comparison of chosen alternatives, but not the identification of new alternatives. The exploration and discussion of new alternatives requires effort and can be a major bottleneck in consensus building and, in our study, it occurred frequently—21.9% of all message class counts were proposing alternatives (Table 1, rows 4 and 5). However, our visual awareness technique provides no direct support for this sub-process of consensus building which may have led to our system's inability to significantly increase overall efficiency.

6.2 Effectiveness

Regarding effectiveness, visual awareness appears to encourage groups to consider the preferences of all members—positively affecting participants' perceptions on their groups' satisfaction. Encouraging the user to consider others' preferences, however, can also push them to conform with their groups despite holding disagreements. Thus, our visual awareness technique could have also negatively impacted satisfaction at an individual level—resulting in the absence of a significant difference in individual satisfaction during our study.

In terms of perceived group-level satisfaction, our technique could increase awareness of preferences and allow participants to more easily consider all of their members' preferences. Thus, participants could have been encouraged to cooperate, either by tuning their own preferences to match their group members' or by bringing up others' preferences during the discussion. Similar to our results, Hong et al. [29] demonstrated that increasing the salience of preferences encourages groups to holistically consider members' preferences. In our study, some groups even leveraged visual awareness to structure their discussions to ensure that all of their members' selection preferences had been addressed. Therefore, by encouraging cooperation and a holistic consideration of members, visual awareness could increase the user's confidence in their groups' satisfaction with the final sequence—increasing perceived effectiveness.

Despite improving perceptions on group-level satisfaction, the visual awareness technique did not significantly increase individual satisfaction towards the final sequence. This contrast in group and individual satisfaction could be attributed to visual awareness allowing groups to easily distinguish disagreeing members and, thus, encouraging the agreeing majority to demand adjustments from the disagreeing member. However, our discourse analysis showed that these demanding messages only accounted for around 1% of all message classes (Table 1, row 9) and did not differ between conditions. Alternatively, it is possible that group members were socially pressured to make changes themselves—without demands from other members—as the technique spotlights sequence changes that reduce disagreements (e.g., the plus button in the list of missing nodes). This could imply groupthink [34]—members hold unresolved disagreements even if superficial consensus is reached—and lower individual satisfaction [43].

6.3 Design Factors and Considerations

The value of our work lies in exploring the nature of CoSeq and introducing a system that takes a first step at supporting said tasks. However, CoSeq tasks are diverse and constituted by various factors that affect the design of effective support. To aid future work in this space, we characterize factors relevant to CoSeq tasks: group size, group hierarchy, sequence length, number of alternative items, and opinion aspects. For each of these factors, we discuss what our work and previous work have identified and suggest considerations for designers aiming to support CoSeq tasks.

Group size: In our work, we explored CoSeq tasks in the context of small groups. We demonstrated that, even in small groups, several challenges inhibit consensus building in CoSeq tasks. For instance, despite groups only having three members, groups had to expend effort to mitigate the "chaotic" [17] or disorganized [28] nature of chat-based discussions—around 8% of all message class counts were of the managing type. In CoSeq tasks with larger groups, this problem will likely be amplified and alternative design choices may be required. For example, Zhang et al. [66] allowed groups of more than 10 crowdworkers to coordinate their sequencing with a to-do list, instead of a discussion. Alternatively, discussions of larger groups can be supported by employing a dedicated mediator [42], or by adopting more organized forms of communication (e.g., threaded discussion forums or scripted chat [16]).

Group hierarchy: Most investigations into CoSeq tasks have focused on groups with relatively flat hierarchies—members hold similar status. Our work focused on groups of friends, and previous work focused on crowdworkers [52, 66] or conference organizers [37]. However, in groups with a leader or large status differences, the opinions of high status members can precede the group's collective opinions. Thus, the support required by these groups may differ and requires further study. However, we believe that the benefit of visual awareness in facilitating the comparison of opinions can still serve a purpose in these hierarchical groups. For example, the system can be adapted to allow all members to participate in individual sequencing, but only allowing high status members to use visual awareness to compare and discuss the sequences. This would enable the high status members to maintain the authority of making the final decision while informing them about their groups' opinions.

Sequence length: We investigated CoSeq tasks which involved short to medium-length sequences (five to seven nodes). With sequences of these lengths, our results showed that visual awareness could hamper efficiency by displaying excessive information. Thus, in tasks with shorter sequences, it can be preferable to not include visual awareness as perceiving preferences would be easier, making the additional interactive components more distracting than beneficial. In scenarios with longer sequences, visual awareness could be designed such that the amount of information presented at a time is manageable and comprehensible. For example, Monroe et al. [47] simplified long health records into shorter sequences—although their support focused on the comparison of a large number of sequences.

Number of alternative items: In our study, groups constructed sequences from a set of 20 alternative items (i.e., POIs). As previously discussed, identifying and discussing new alternatives could be a major bottleneck in CoSeq tasks. In scenarios with small to medium-sized set of alternatives, support for the identification of new alternative selections or orderings could overcome this bottleneck. For example, natural language processing could be used to analyze a group's discussion and suggest alternatives. However, if there are hundreds or thousands of alternative items, concurrently exploring selections and orderings may be inefficient. For instance, Kim et al. [37] investigated the scheduling of hundreds of papers for a conference. Thus, their support was designed to allow organizers to first allocate papers into sessions (i.e., selection) and, after allocation, organizers discussed to decide on the order of papers within the sessions.

Opinion aspects (e.g., preference, knowledge, or expertise): In our work, we focused on a CoSeq task in which group members' opinions are mainly based on their subjective preferences. We controlled for other aspects, such as knowledge or expertise, by ensuring that none of the participants had previously visited the cities presented in the study. However, in more serious CoSeq domains, such as managing software teams' workflows or scheduling admissions to intensive care units, the integration of individual members' expertise and knowledge may be more crucial than considering preferences. Thus, when designing support for a CoSeq task, it is critical to consider which aspects of opinions are relevant to the domain and how to support awareness of these—e.g., allow members to specify their expertise and visualize this in relation to sequences.

7 LIMITATIONS

Our studies allowed us to acquire a deeper understanding of the design of visual awareness to support consensus building in CoSeq, and the effect of this type of support on the efficiency and effectiveness of the said process. However, we acknowledge that our work has the following limitations:

- **Group type**: We conducted our study with groups of friends to control for group dynamics. In our summative study, the average total initial edit distance for all the sessions was 23.4 (SD=4.52), which indicates that participants in the same group had relatively different preferences and that they engaged in conflicts. However, compared to other types of groups (e.g., strangers, co-workers, or family members), groups of friends are more likely to be similar in personality, preferences, and be more close to each other. These similarities could have helped participant groups reach a consensus with more ease and, thus, affected the generalizability of our results. To gain a more comprehensive understanding, further work could investigate CoSeq tasks with diverse types of groups or while varying the degree of similarities in personality between group members.
- **Group size**: We conducted our study with groups of three. Since groups had an odd number of members, it is possible that majority control occurred in some groups—i.e., two members agree and the remaining member is pressured to agree. While our discourse analysis showed that participants rarely demanded others to make sequence changes and majority control could also occur in even numbered groups, further work with diverse group sizes (e.g., larger or even-numbered) is needed to verify the effect of this phenomenon in CoSeq tasks.
- **Chosen CoSeq domain**: Although we believe that our findings and implications are applicable to a variety of CoSeq task types, our studies focused on travel itinerary planning. We acknowledge that the characteristics of travel planning, such as the geographical distances between POIs, may have affected the consensus building processes we observed. Although other CoSeq tasks may also have dependencies (e.g., topics which are prerequisites to others in course scheduling), further work is needed to validate the generalizability of our findings.
- **Task complexity**: In our studies, participants only considered sequences including five to seven nodes. The design considerations of visual awareness and its effect may differ in more complex tasks.
- **Controlled setting**: Our summative study was designed as a controlled lab experiment to concentrate on understanding the effects of visual awareness. However, the study does not fully represent real-world situations (e.g., asynchronous and remote settings). Future work could explore the effects of visual awareness in CoSeq in diverse settings through a deployment study.

8 CONCLUSION

In this paper, we illustrate the challenges of consensus building in CoSeq tasks and provide insights into how to support the process. We conducted a formative study through which we extracted a set of design requirements for supporting visual awareness in CoSeq. Based on these design requirements, we presented a novel technique and instantiated it in Twine, a collaborative travel itinerary planning system. Leveraging the system as a research apparatus, we presented findings from a summative study. We found that visual awareness can increase the efficiency of consensus building by facilitating the communication of preferences, and the identification and resolution of conflicts. Additionally, we found from participants' responses that visual awareness can also implicitly structure discussions to allow for fair participation, and encourage group members to be more cooperative and considerate. We anticipate that our findings and our visual awareness technique can guide and inspire the design of future applications to support a wide variety of CoSeq tasks.

9 ACKNOWLEDGMENTS

This research was supported by the KAIST UP Program. The authors would like to thank the members of KIXLAB for their constant support and feedback. We also thank our study participants for their time and engagement with the task. Finally, we thank the reviewers for helping us improve our paper through their feedback in each iteration stage of the review process.

REFERENCES

[1] Christopher Ahlberg and Ben Shneiderman. 2003. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. In *The Craft of Information Visualization*, Benjamin B. Bederson and Ben Shneiderman (Eds.). Morgan Kaufmann, San Francisco, 7–13. https://doi.org/10.1016/B978-155860915-0/50004-4

[2] S. Alonso, E. Herrera-Viedma, F. J. Cabrerizo, F. Chiclana, and F. Herrera. 2007. Visualizing Consensus in Group Decision Making Situations. In *2007 IEEE International Fuzzy Systems Conference*. 1–6. https://doi.org/10.1109/FUZZY.2007.4295642

[3] Basak Alper, Benjamin Bach, Nathalie Henry Riche, Tobias Isenberg, and Jean-Daniel Fekete. 2013. Weighted Graph Comparison Techniques for Brain Connectivity Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 483–492. https://doi.org/10.1145/2470654.2470724

[4] K. Andrews, M. Wohlfahrt, and G. Wurzinger. 2009. Visual Graph Comparison. In *2009 13th International Conference Information Visualisation*. 62–67. https://doi.org/10.1109/IV.2009.108

[5] Michael Avery, Barbara Stribel, Lonnie Weiss, and Brian Auvine. 1981. *Building United Judgment: A Handbook for Consensus Decision Making*. The Center for Conflict Resolution, Madison.

[6] Jordan B. Barlow. 2013. Emergent Roles in Decision-Making Tasks Using Group Chat. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 1505–1514. https://doi.org/10.1145/2441776.2441948

[7] Philip Bille. 2005. A survey on tree edit distance and related problems. *Theoretical Computer Science* 337, 1 (2005), 217–239. https://doi.org/10.1016/j.tcs.2004.12.030

[8] Robert O Briggs, Gwendolyn L Kolfschoten, and Gert-Jan de Vreede. 2005. Toward a theoretical model of consensus building. *AMCIS 2005 Proceedings* (2005), 12.

[9] Moira Burke and Robert Kraut. 2008. Mopping up: Modeling Wikipedia Promotion Decisions. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (San Diego, CA, USA) *(CSCW '08)*. ACM, New York, NY, USA, 27–36. https://doi.org/10.1145/1460563.1460571

[10] CT Lawrence Butler and Amy Rothstein. 2007. *On conflict and consensus: A handbook on formal consensus decisionmaking*. Citeseer.

[11] Munmun De Choudhury, Moran Feldman, Sihem Amer-Yahia, Nadav Golbandi, Ronny Lempel, and Cong Yu. 2010. Automatic Construction of Travel Itineraries Using Social Breadcrumbs. In *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia* (Toronto, Ontario, Canada) *(HT '10)*. ACM, New York, NY, USA, 35–44. https://doi.org/10.1145/1810617.1810626

[12] Paul Dourish and Matthew Chalmers. 1994. Running out of Space: Models of Information Navigation.. In *Proceedings of HCI '94*. ACM Press, Glasgow, Scotland.

[13] Fan Du, Sana Malik, Georgios Theocharous, and Eunyee Koh. 2018. *Personalizable and Interactive Sequence Recommender System*. ACM, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188506

[14] Fan Du, Catherine Plaisant, Neil Spring, Kenyon Crowley, and Ben Shneiderman. 2019. EventAction: A Visual Analytics Approach to Explainable Recommendation for Event Sequences. *ACM Trans. Interact. Intell. Syst.* 9, 4, Article 21, 31 pages. https://doi.org/10.1145/3301402

[15] Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. 2017. Finding Similar People to Guide Life Choices: Challenge, Design, and Evaluation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 5498–5544. https://doi.org/10.1145/3025453.3025777

[16] Shelly Farnham, Harry R. Chesley, Debbie E. McGhee, Reena Kawal, and Jennifer Landau. 2000. Structured Online Interactions: Improving the Decision-Making of Small Discussion Groups. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (Philadelphia, Pennsylvania, USA) *(CSCW '00)*. ACM, New York, NY, USA, 299–308. https://doi.org/10.1145/358916.359001

[17] Fiona E. Fox, Marianne Morris, and Nichola Rumsey. 2007. Doing Synchronous Online Focus Groups With Young People: Methodological Reflections. *Qualitative Health Research* 17, 4 (2007), 539–547. https://doi.org/10.1177/1049732306298754 arXiv:https://doi.org/10.1177/1049732306298754 PMID: 17416707.

[18] L. Alberto Franco, Etiënne A.J.A. Rouwette, and Hubert Korzilius. 2016. Different paths to consensus? The impact of need for closure on model-supported group conflict management. *European Journal of Operational Research* 249, 3 (2016), 878–889. https://doi.org/10.1016/j.ejor.2015.06.056

[19] Xinbo Gao, Bing Xiao, Dacheng Tao, and Xuelong Li. 2010. A survey of graph edit distance. *Pattern Analysis and Applications* 13, 1 (2010), 113–129. https://doi.org/10.1007/s10044-008-0141-y

[20] M. Gleicher. 2018. Considerations for Visualizing Comparison. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 413–423. https://doi.org/10.1109/TVCG.2017.2744199

[21] Nitesh Goyal and Susan R. Fussell. 2016. Effects of Sensemaking Translucence on Distributed Collaborative Analysis. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. ACM, New York, NY, USA, 288–302. https://doi.org/10.1145/2818048.2820071

[22] Nitesh Goyal, Gilly Leshed, Dan Cosley, and Susan R. Fussell. 2014. Effects of Implicit Sharing in Collaborative Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 129–138. https://doi.org/10.1145/2556288.2557229

[23] Carl Gutwin and Saul Greenberg. 1998. Design for Individuals, Design for Groups: Tradeoffs between Power and Workspace Awareness. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '98)*. ACM, New York, NY, USA, 207–216. https://doi.org/10.1145/289444.289495

[24] A. H. Hajizadeh, M. Tory, and R. Leung. 2013. Supporting Awareness through Collaborative Brushing and Linking of Tabular Data. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2189–2197. https://doi.org/10.1109/TVCG.2013.197

[25] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[26] Jeffrey Heer and Maneesh Agrawala. 2008. Design Considerations for Collaborative Visual Analytics. *Information Visualization* 7, 1 (2008), 49–62. https://doi.org/10.1057/palgrave.ivs.9500167 arXiv:https://doi.org/10.1057/palgrave.ivs.9500167

[27] Jeffrey Heer, Fernanda B. Viégas, and Martin Wattenberg. 2007. Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. ACM, New York, NY, USA, 1029–1038. https://doi.org/10.1145/1240624.1240781

[28] Starr Roxanne Hiltz, Kenneth Johnson, and Murray Turoff. 1986. Experiments in Group Decision Making Communication Process and Outcome in Face-to-Face Versus Computerized Conferences. *Human Communication Research* 13, 2 (1986), 225–252. https://doi.org/10.1111/j.1468-2958.1986.tb00104.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2958.1986.tb00104.x

[29] Sungsoo (Ray) Hong, Minhyang (Mia) Suh, Nathalie Henry Riche, Jooyoung Lee, Juho Kim, and Mark Zachry. 2018. Collaborative Dynamic Queries: Supporting Distributed Small Group Decision-Making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173640

[30] Sungsoo (Ray) Hong, Minhyang (Mia) Suh, Tae Soo Kim, Irina Smoke, Sangwha Sien, Janet Ng, Mark Zachry, and Juho Kim. 2019. Design for Collaborative Information-Seeking: Understanding User Challenges and Deploying Collaborative Dynamic Queries. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 106 (Nov. 2019), 24 pages. https://doi.org/10.1145/3359208

[31] Hsun-Ping Hsieh, Cheng-Te Li, and Shou-De Lin. 2012. TripRec: Recommending Trip Routes from Large Scale Check-in Data. In *Proceedings of the 21st International Conference on World Wide Web* (Lyon, France) *(WWW '12 Companion)*. ACM, New York, NY, USA, 529–530. https://doi.org/10.1145/2187980.2188111

[32] Petra Isenberg and Danyel Fisher. 2009. Collaborative Brushing and Linking for Co-located Visual Analytics of Document Collections. *Computer Graphics Forum* 28, 3 (2009), 1031–1038. https://doi.org/10.1111/j.1467-8659.2009.01444.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2009.01444.x

[33] Hiroko Itakura. 2001. Describing conversational dominance. *Journal of Pragmatics* 33, 12 (2001), 1859–1880. https://doi.org/10.1016/S0378-2166(00)00082-5

[34] Irving L Janis. 1972. *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes.* Houghton Mifflin, Oxford, England. viii, 277–viii, 277 pages.

[35] Ruogu Kang, Aimee Kane, and Sara Kiesler. 2014. Teammate Inaccuracy Blindness: When Information Sharing Tools Hinder Collaborative Analysis. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. ACM, New York, NY, USA, 797–806. https://doi.org/10.1145/2531602.2531681

[36] Gavin J. Kilduff, Robb Willer, and Cameron Anderson. 2016. Hierarchy and Its Discontents: Status Disagreement Leads to Withdrawal of Contribution and Lower Group Performance. *Organization Science* 27, 2 (2016), 373–390. https://doi.org/10.1287/orsc.2016.1058 arXiv:https://doi.org/10.1287/orsc.2016.1058

[37] Juho Kim, Haoqi Zhang, Paul André, Lydia B. Chilton, Wendy Mackay, Michel Beaudouin-Lafon, Robert C. Miller, and Steven P. Dow. 2013. Cobi: A Community-Informed Conference Scheduling Tool. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) *(UIST '13)*. ACM, New York, NY, USA, 173–182. https://doi.org/10.1145/2501988.2502034

[38] Tae Soo Kim, Sungsoo (Ray) Hong, Nitesh Goyal, Jeongyeon Kim, and Juho Kim. 2020. Consensus Building in Collaborative Sequencing with Visual Awareness. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. ACM, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3382906

[39] Aniket Kittur and Robert E. Kraut. 2010. Beyond Wikipedia: Coordination and Conflict in Online Production Groups. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, Georgia, USA) *(CSCW '10)*. ACM, New York, NY, USA, 215–224. https://doi.org/10.1145/1718918.1718959

[40] Nicholas Kong, Tovi Grossman, Björn Hartmann, Maneesh Agrawala, and George Fitzmaurice. 2012. Delta: A Tool for Representing and Comparing Workflows. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. ACM, New York, NY, USA, 1027–1036. https://doi.org/10.1145/2207676.2208549

[41] Bibb Latané, Kipling Williams, and Stephen Harkins. 1979. Many hands make light the work: The causes and consequences of social loafing. *Journal of personality and social psychology* 37, 6 (1979), 822.

[42] 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. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376609

[43] 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

[44] Sana Malik and Eunyee Koh. 2016. High-Volume Hypothesis Testing for Large-Scale Web Log Analysis. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. ACM, New York, NY, USA, 1583–1590. https://doi.org/10.1145/2851581.2892487

[45] Joseph Edward McGrath. 1984. *Groups: Interaction and performance*. Vol. 14. Prentice-Hall Englewood Cliffs, NJ.

[46] Andreas Mojzisch and Stefan Schulz-Hardt. 2010. Knowing Others' Preferences Degrades the Quality of Group Decisions. *Journal of personality and social psychology* 98 (05 2010), 794–808. https://doi.org/10.1037/a0017627

[47] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman. 2013. Temporal Event Sequence Simplification. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2227–2236. https://doi.org/10.1109/TVCG.2013.200

[48] Masato Nomiyama, Toshiki Takeuchi, Hiroyuki Onimaru, Tomohiro Tanikawa, Takuji Narumi, and Michitaka Hirose. 2018. Xnavi: Travel Planning System Based on Experience Flows. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 27 (March 2018), 25 pages. https://doi.org/10.1145/3191759

[49] Babajide Osatuyi, Starr Roxanne Hiltz, and Katia Passerini. 2016. Seeing is believing (or at least changing your mind): The influence of visibility and task complexity on preference changes in computer-supported team decision making. *Journal of the Association for Information Science and Technology* 67, 9 (2016), 2090–2104. https://doi.org/10.1002/asi.23555 arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23555

[50] Ei Pa Pa Pe-Than, Laura Dabbish, and James D. Herbsleb. 2018. Collaborative Writing on GitHub: A Case Study of a Book Project. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Jersey City, NJ, USA) *(CSCW '18)*. ACM, New York, NY, USA, 305–308. https://doi.org/10.1145/3272973.3274083

[51] Chih-Hung Peng, Nicholas H. Lurie, and Sandra A. Slaughter. 2019. Using Technology to Persuade: Visual Representation Technologies and Consensus Seeking in Virtual Teams. *Information Systems Research* 30, 3 (2019), 948–962. https://doi.org/10.1287/isre.2019.0843 arXiv:https://doi.org/10.1287/isre.2019.0843

[52] Joseph Rafidi. 2013. Real-Time Trip Planning with the Crowd. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. ACM, New York, NY, USA, 2743–2748. https://doi.org/10.1145/2468356.2479503

[53] Johnny Saldaña. 2009. *The coding manual for qualitative researchers.* Sage Publications Ltd, Thousand Oaks, CA. xi, 223–xi, 223 pages.

[54] B. Shneiderman. 1996. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*. 336–343. https://doi.org/10.1109/VL.1996.545307

[55] Nonnadda Silamai, Narongchai Khamchuen, and Santi Phithakkitnukoon. 2017. TripRec: Trip Plan Recommendation System That Enhances Hotel Services. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Maui, Hawaii) *(UbiComp '17)*. ACM, New York, NY, USA, 412–420. https://doi.org/10.1145/3123024.3124414

[56] Martin Stettinger, Alexander Felfernig, Gerhard Leitner, and Stefan Reiterer. 2015. Counteracting Anchoring Effects in Group Decision Making. In *User Modeling, Adaptation and Personalization*, Francesco Ricci, Kalina Bontcheva, Owen Conlan, and Séamus Lawless (Eds.). Springer International Publishing, Cham, 118–130.

[57] Lawrence Susskind, Sarah McKearnan, and Jennifer Thomas-Larmer. 1999. *The Consensus Building Handbook: A Comprehensive Guide to Reaching Agreement.* Sage Publications, Thousand Oaks, California. https://doi.org/10.4135/9781452231389

[58] Kendall Taylor, Kwan Hui Lim, and Jeffrey Chan. 2018. Travel Itinerary Recommendations with Must-See Points-of-Interest. In *Companion Proceedings of the The Web Conference 2018* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1198–1205. https://doi.org/10.1145/3184558.3191558

[59] Atima Tharatipyakul, Hyowon Lee, Shengdong Zhao, and Richard C. Davis. 2016. Supporting the Comparison of Alternative Stories. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (Launceston, Tasmania, Australia) *(OzCHI '16)*. ACM, New York, NY, USA, 266–270. https://doi.org/10.1145/3010915.3010963

[60] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 4157 (1974), 1124–1131. https://doi.org/10.1126/science.185.4157.1124 arXiv:https://science.sciencemag.org/content/185/4157/1124.full.pdf

[61] F. B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. 2007. ManyEyes: a Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1121–1128. https://doi.org/10.1109/TVCG.2007.70577

[62] W. Willett, J. Heer, and M. Agrawala. 2007. Scented Widgets: Improving Navigation Cues with Embedded Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1129–1136. https://doi.org/10.1109/TVCG.2007.70589

[63] Krist Wongsuphasawat, John Alexis Guerra Gómez, Catherine Plaisant, Taowei David Wang, Meirav Taieb-Maimon, and Ben Shneiderman. 2011. LifeFlow: Visualizing an Overview of Event Sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 1747–1756. https://doi.org/10.1145/1978942.1979196

[64] Haixia Xu and Libby V Morris. 2007. Collaborative course development for online courses. *Innovative Higher Education* 32, 1 (2007), 35.

[65] Fatemeh Zahedi. 1986. Group Consensus Function Estimation When Preferences are Uncertain. *Operations Research* 34, 6 (1986), 883–894. https://doi.org/10.1287/opre.34.6.883 arXiv:https://doi.org/10.1287/opre.34.6.883

[66] Haoqi Zhang, Edith Law, Rob Miller, Krzysztof Gajos, David Parkes, and Eric Horvitz. 2012. Human Computation Tasks with Global Constraints. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. ACM, New York, NY, USA, 217–226. https://doi.org/10.1145/2207676.2207708

[67] Lanyun Zhang and Xu Sun. 2016. Designing a Trip Planner Application for Groups: Exploring Group Tourists? Trip Planning Requirements. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. ACM, New York, NY, USA, 1329–1336. https://doi.org/10.1145/2851581.2892301

[68] Jian Zhao, Zhicheng Liu, Mira Dontcheva, Aaron Hertzmann, and Alan Wilson. 2015. MatrixWave: Visual Comparison of Event Sequence Data. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, 259–268. https://doi.org/10.1145/2702123.2702419

[69] Roshanak Zilouchian Moghaddam, Zane Nicholson, and Brian P. Bailey. 2015. Procid: Bridging Consensus Building Theory with the Practice of Distributed Design Discussions. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. ACM, New York, NY, USA, 686–699. https://doi.org/10.1145/2675133.2675272

Received June 2020; revised October 2020; accepted December 2020