ConsensUs: Supporting Multi-Criteria Group Decisions by Visualizing Points of Disagreement

WEICHEN LIU, University of California San Diego
SIJIA XIAO, Georgia Institute of Technology
JACOB T. BROWNE, University of California San Diego
MING YANG, Cornell University
STEVEN P. DOW, University of California San Diego

Groups often face difficulty reaching consensus. For complex decisions with multiple criteria, verbal and written discourse alone may impede groups from pinpointing and moving past fundamental disagreements. To help support consensus building, we introduce ConsensUs, a novel visualization tool that highlights disagreement by asking group members to quantify their subjective opinions across multiple criteria. To evaluate this approach, we conducted a between-subjects experiment with 87 participants on a comparative hiring task. The study compared three modes of sensemaking on a group decision: written discourse only, visualization only, and written discourse plus visualization. We confirmed that the visualization helped participants identify disagreements within the group and then measured subsequent changes to their individual opinions. The results show that disagreement highlighting led participants to align their ratings more with the opinions of other group members. While disagreement highlighting led to better score alignment, participants reported a number of reasons for shifting their score, from genuine consensus to appeasement. We discuss further research angles to understand how disagreement highlighting affects social processes and whether it produces objectively better decisions.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing systems and tools**; **Visualization systems and tools**; *Interactive systems and tools*;

Additional Key Words and Phrases: Multi-criteria decisions, group decision making, consensus building, information visualization

Authors' addresses: W. Liu, 1415B 19th AVE, Seattle WA 98122; email: liuweichen.xyz@gmail.com; S. Xiao, #1218, 1000 Northside St NW, Atlanta, GA 30318; email: xiaosijia@gatech.edu; J. T. Browne, 4430 Campus Avenue, Apt 9, San Diego, CA, 92116; email: jacobbrowne8@gmail.com; M. Yang, 41-42 24th Street, APT 508, Long Island City, NY, 11101; email: my434@cornell.edu; S. P. Dow, 9500 Gilman Dr. La Jolla, CA 92093; email: spdow@ucsd.edu.

1 INTRODUCTION

Groups are often charged with reaching consensus on important decisions. Families must agree on which house or car to buy; hiring committees must decide on which potential employee to hire; local government boards choose between different investments in their community. Such decisions often involve comparing multiple alternatives along one or more criteria, a topic known in the literature as multi-criteria decision making (MCDM) (Gal et al. 2013).

For particularly complex decisions, groups delineate different criteria to help organize information about alternatives (Gal et al. 2013). Group members offer multiple perspectives and differing opinions on how to weigh criteria relative to each other (Herrera et al. 1996). Typically, groups communicate about such decisions through verbal discourse or through computer-mediated chat or email (McGrath and Hollingshead 1990; Haskins and Nilssen 2015). However, verbal discussions can be dominated by one individual, either by means of conversation dominance (Itakura 2001) or social status (Dubrovsky et al. 1991), which may leave groups vulnerable to groupthink (Janis 1972), production blocking (Lamm and Trommsdorff 1973), or social loafing (Latane et al. 1979). Group decisions can also be influenced by the first advocate for a particular position (Weisband 1992), by group polarization (decisions initially preferred by a plurality of members) (Myers and Lamm 1976; Dennis et al. 1997), or by group biases (Jones and Roelofsma 2000; Montibeller and Winterfeldt 2015; Mannion and Thompson 2014). Individuals can be "anchored" to their initial impression (Tversky and Kahneman 1975), or conversely, could exhibit false consensus by avoiding or ignoring conflict within a group (Franco et al. 2016).

Researchers have explored how technology could mediate group decisions as a way to overcome barriers in face-to-face discourse (Benbunan-Fich et al. 2003; Schmidt et al. 2001). Tools for model-supported conflict management have sought to reconcile group differences, for example, by promoting value-focused thinking (Sheng et al. 2005; Keeney 1996), ensuring information sharing (Lam and Schaubroeck 2000), fostering a positive tone (Zilouchian Moghaddam et al. 2015), or by structuring group discussion (Farnham et al. 2000). However, computer-mediated discussions are also known to increase delays, create more outspoken advocacy, decrease member satisfaction, and lead to riskier decisions (Kiesler and Sproull 1992; Hiltz et al. 1986; Baltes et al. 2002).

Our research explores how technology—specifically how an interactive visualization—affects group consensus building above and beyond traditional methods of discourse. Effective visualization can help users discover actionable insights from data and grapple with complex problems (Yi et al. 2007). For example, ValueCharts (Carenini and Loyd 2004) uses visualization to help users make preferential choices in MCDM scenarios. Taheri et al. (2015) argue that visualization could help identify key sticking-points and highlight participants with influential opinions. Our research explores how this kind of visual support affects collaborative decisions.

In this article, we build on this prior work to introduce a visualization approach for multi-criteria group decision making based on consensus building theory. Briggs et al. (2005) articulates consensus building as a social process: first a group member makes a "proposal" and the group evaluates the extent they are willing to commit to that proposal. If all members are willing to commit, then consensus has been achieved. If some members are not willing to commit to the proposal, then a group must diagnose the conflict using various strategies. Challenges often revolve around a group's ability to *identify the causes of disagreement* (Jehn 1997) and to *form a holistic view of the group's perspective* (Avery et al. 1981).

We designed and built a real-time interactive visualization tool for MCDM called ConsensUs. In the tool, group members provide independent numeric ratings along key criteria (see Figure 1) and articulate opinions with written comments. After members provide initial independent opinions, they see a group page that contains a rank-order visualization of the group average ratings for each alternative. The interface explicitly highlights the amount of disagreement among members and

the distance between each group member's rating and the group average along each criteria (see Figure 2). It also breaks down ratings for each criteria and again highlights group disagreements for each criterion. Members can drill down further to find and compare different group members' opinions. Members can update their arguments and ratings in real time.

Our evaluation focused on the relative benefits of disagreement highlighting as a supplement to written arguments. We conducted a between-subjects experiment with 87 participants taking part in a mock admissions committee for an engineering program. Participants read about three fictional candidates and provided an independent assessment by writing an argument and recording a rating for each candidate along predetermined decision criteria. A set of six confederate voters were randomly selected from a pre-populated subject pool of people who performed only the individual task. As the key independent variable, participants were presented one of three different representations of the other committee members' opinions: in the *Arguments-only* condition, participants only see written arguments by other committee members. In the *Visualization-only* condition, participants viewed the ConsensUs' visualization of the subjective ratings of other committee members with key disagreements highlighted. In the *Both* condition, participants saw both the visualization and the written arguments. As a manipulation check, we asked participants questions about the underlying group disagreements. Then, they had an opportunity to update their candidate ratings, which provided a measure of preference change after viewing group information. Participants also filled out a survey regarding their perceived ability to identify disagreements and their rationale for their final decision.

The results show that disagreement highlighting helped participants articulate the key points of group disagreement (confirming our manipulation check), and subsequently led participants to change their opinion more toward the group's opinion. The survey revealed this happened for a variety of reasons, from genuine consensus building to shifting their ratings to simply appear more aligned with the group. Further qualitative analysis indicates that participants benefit both from the visualization support to locate points of disagreement and written discourse to understand others' rationale.

This research offers the key design insight that highlighting points of disagreement in group decisions can affect individual scoring behavior. While participants express a range of reasons for shifting their scores, from authentic consensus to appeasement, they ultimately end up with better group alignment. This article contributes (1) a research review of key factors and existing group-support software designed to improve consensus building in decision making, (2) a novel interactive visualization tool for supporting groups informed by consensus-building theory, (3) an empirical study that investigates the relative value of visualizing group disagreements and reveals how this leads to better score alignment, and (4) an open-coding analysis that explores why people shift their scores.

2 RELATED WORK

By way of background, we review the foundational literature on identifying conflicts and building consensus for decision making. We also summarize a number of prior systems designed to support consensus building and decision making in various contexts.

2.1 Identifying Conflict and Building Consensus

Consensus building in groups can be defined as the process of coming to a decision in which all group members find acceptable, despite objections that may still be present (Avery et al. 1981). When successful, this process articulates key aspects of the problem, results in consideration of different proposals, and fortifies commitment to the decision (Susskind et al. 1999). Briggs et al. (2005) offer a theoretical model of consensus building, where members share and discuss

proposals, diagnose conflicts, and use strategies to bridge disagreements, typically by uncovering the underlying rationale for members' opinions.

Group members address conflict differently. Some members avoid voicing dissent to help maintain internal social relationships in work groups (Hackman and Kaplan 1974; Myers and Lamm 1976). People may influence one another so much that they ignore their own concerns and only focus on the publicly stated arguments by others (Sunstein 2006). Failing to surface disagreements may lead to "false consensus," where group members overestimate the extent to which their opinions align with others (Mullen et al. 1985; Marks and Miller 1987). Assuming conflicts do surface, groups may still choose "win-lose" solutions that fail to address the concerns of all members (Franco et al. 2016). Groups typically use verbal discourse to uncover whether disagreeing members have different mental models, information, goals, meaning and/or taste (Briggs et al. 2005), which can then be negotiated to reach resolution.

Individual biases can also negatively affect the consensus-building process. For example, individual members may be subject to an anchoring effect where people are attached to their initial opinion and unmoved by a group's opinion (Tversky and Kahneman 1975). Individuals are also known to exhibit confirmation bias where they resist any disconfirming evidence to their initial preferences (Wason 1960; Toma and Butera 2009; Nickerson 1998). Confirmation bias can hamper group decisions when group member's initial preferences dominate (Greitemeyer and Schulz-Hardt 2003).

Procedural strategies can help forge a more sophisticated consensus-building process. For example, Stettinger et al. argues that anchoring in group decision processes can be reduced if a group's preferences are revealed only after all group members have articulated their individual preferences (Avery et al. 1981; Susskind et al. 1999). Briggs et al. (2005) argue for a consensus-building process that establishes criteria by which to compare alternatives, externalizes points of agreement and disagreement to help scaffold the discussion, ensures equal participation, and maintains a holistic view of the group's opinion (Briggs et al. 2005). Our research utilizes consensus-building theory to design a tool that surfaces and facilitates discussion around group disagreements.

2.2 Technology to Support Group Decision Making

Many researchers have created group support systems that attempt to make group processes more salient through information visualization. In particular, numerous technologies have emerged to support multi-criteria comparisons (Chimera 1992; Herrera et al. 1996; Chen 2000; Keim et al. 2002; Andrienko and Andrienko 2002; Romero et al. 2012; Gratzl et al. 2013; Aseniero et al. 2015; Wittenburg and Turchi 2016; Carenini and Loyd 2004; Conati et al. 2014; Gratzl et al. 2013; Yi 2008). Most similar to our work is ValueCharts (Carenini and Loyd 2004), a visualization tool for multi-criteria decision making designed for a single user. Group ValueCharts extended this work with the goal of increasing group participation, transparency, and comprehensibility (Bajracharya 2014), but this work does not focus on explicitly externalizing points of group disagreement to scaffold discussions.

Other researchers have created systems aimed specifically at supporting consensus building in different contexts. For example, Moghaddam et al. created Procid to help designers visually organize discussions around proposals, define and rate against criteria, and indicate support (Zilouchian Moghaddam et al. 2015). Procid also analyzes the sentiment of comments and guides users toward providing more positive feedback on proposals. Alonso et al. explored how to visualize consensus among a distributed group by representing the "distance" between each voter as an indication of the closeness of opinions (Alonso et al. 2007a, 2007b). Kriplean et al. (2007) emphasized the need for conflict identification for discussion pages on Wikipedia where editors would need to reach consensus on an article's content. Osatuyi et al. (2015) found that making

individual preferences visible in the context of a hiring and firing task helped lead participants to change their preferences to align more with the group opinion. Further, large-scale argumentation systems, such as the Deliberatorium (Klein 2012), build on argumentation theory to structure large complex argument maps.

Our work is particularly focused on how we might leverage techniques in information visualization to scaffold group problem solving. Visualization can help decision-makers discover hidden patterns, understand relationships, generate and evaluate hypotheses, and alleviate uncertainty (Taheri 2015). For example, in collaborative problem solving, Balakrishnan et al. (2008) demonstrated that presenting a visualization with pertinent information and shared interaction positively augments collaboration. LineUp uses bar charts to compare alternatives among multiple attributes (Gratzl et al. 2013). Gratzl et al. show that visualization techniques can help people successfully solve complex ranking tasks—where alternatives are placed in order of importance—in a shorter amount of time. Aseniero et al. (2015) created a hybrid visualization combining Sankey diagrams and parallel coordinates in a tree view to support decision making for software release planning (Riehmann et al. 2005). OpinionSpace shows a graphical "map" to highlight the most insightful comments in an online discussion; the research shows that such a visualization approach scales better and addresses polarization better than simply listing the information (Faridani et al. 2010).

In a similar vein, a number of tools attempt to create a structured decision process where visualization plays a prominent role (Rosenberg 2015; UNU 2017; Loomio 2017; Polleverywhere 2017). For example, in the Delphi Method, experts independently provide quantifiable views of forecasting and policy decisions and then view visualizations of other experts' opinions in a reflective, iterative process (Gordon 2009; Linstone et al. 1975; Dalkey et al. 1969; Okoli and Pawlowski 2004).

We explore an opportunity here for new research on visualizing opinions for group decisions. Our work explicitly builds on consensus building theory and focuses on highlighting the sources of disagreement as a strategy for convergence. Group members can submit written arguments, as well as ratings that quantify their subjective opinions along multiple criteria. The tool then visualizes the ratings of all group members in a shared representation and highlights the disagreements.

3 CONSENSUS: A TOOL FOR MULTI-CRITERIA GROUP DECISIONS

This section outlines design considerations and features for tools to support MCDM by visualizing subjective data about complex decisions. We built ConsensUs as a Web browser-based system, allowing groups both online and in-person to interact with the tool in real time. We built the front-end visualization with JavaScript and D3, and leveraged Meteor as the full-stack framework. The long-term goal of ConsensUs is to support groups in a real-time consensus-building process.

3.1 Design Guidelines

Building on the prior literature, we outline the main design considerations in this section. The most novel consideration for the design of ConsensUs is to visualize points of disagreements. While most group support systems aggregate information in some way or another, few of them pay specific attention toward highlighting points of disagreement. In theory, externalizing points of agreement and disagreement can help scaffold the discussion and help a group diagnose conflicts (Briggs et al. 2005).

Several other considerations influenced our design of ConsensUs. Based on prior work (Avery et al. 1981; Susskind et al. 1999; Osatuyi et al. 2015; Stettinger et al. 2015), our system asks participants to first provide independent opinions before showing the group overview. Within the independent voting interface, participants can rank all candidates simultaneously in one view, rather than rating them independently or only selecting a preferred option (Hollingshead 1996). In the group interface, ConsensUs provides an overview that allows group members to easily consider
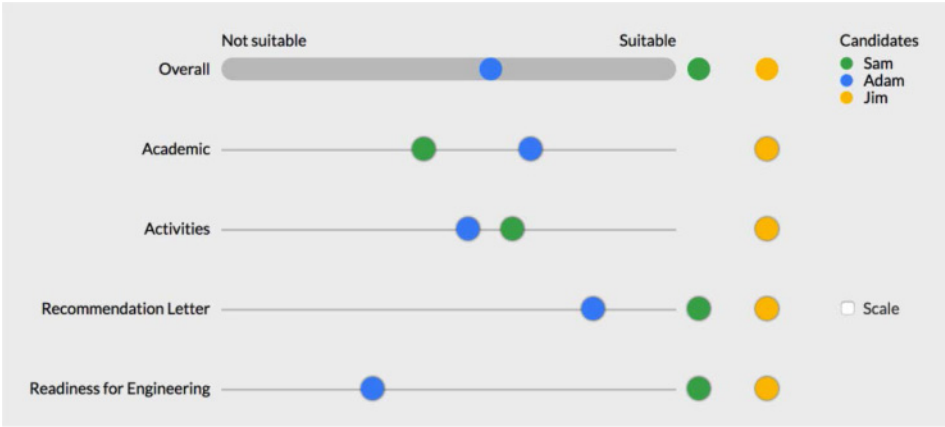
Fig. 1. The ConsensUs independent opinion interface where users rank alternatives relative to each other on a number of criteria. In this view, the user has finished ranking Adam and has started to rank Sam.

the different criteria and alternatives without having to switch between different views (Shneiderman 1996; Aseniero et al. 2015). Moreover, to insure that each participant feels valued and considered in the decision-making process, we designed the system to give all decision-makers equal representation and ability to influence the group average opinion. In other words, the system does not add weights to any members' ratings, although this could be explored in future work.

3.2 Tool Features

ConsensUs structures group decisions around these two key phases: (1) capturing independent opinions and (2) representing group opinions.

3.2.1 Capturing Independent Opinions. In ConsensUs's independent opinion interface, users rank alternatives relative to each other on a number of criteria. As depicted in Figure 1, users can click and drag the colored circles representing the different alternatives on to each criterion line. The criteria are organized as rows and can be customized to list as many criterion as the decision merits. Users can also customize the labels on each criterion spectrum (e.g., by default, we label one side as "not suitable" and the opposite side as "suitable"). The goal was to create a simple and intuitive visual representation for ranking alternatives.

The topmost row reflects the total average rating for each alternative in ranked fashion. This bar updates automatically as users change their relative rankings for different criteria. By default, all criteria contribute equally to the average ratings for each alternative. If users prefer a more quantitative view, they can select the "Scale" checkbox to show a 0 to 10 integer scale along each criterion line. Below the relative ranking area, users can input written arguments to support their position.

3.2.2 Representing Group Opinions. In ConsensUs's group visualization interface, users see an aggregation of all group members' opinions from the individual interface (see Figure 2). The interface includes both written arguments (see Figure 2(e)) and the average ratings regarding the group's preferences (see Figure 2(a)). The interface is laid out similarly to the independent opinion interface such that the criteria remain in the same place and the different colored dots represent the same alternatives. In the group interface, the bigger dots represent the average ratings of the group (including the user) and the smaller dots represent the user's individual ratings.

If a group narrows down their decision to fewer alternatives, they can hide alternatives from the visualization by clicking the corresponding checkbox in the legend. This reduces visual clutter
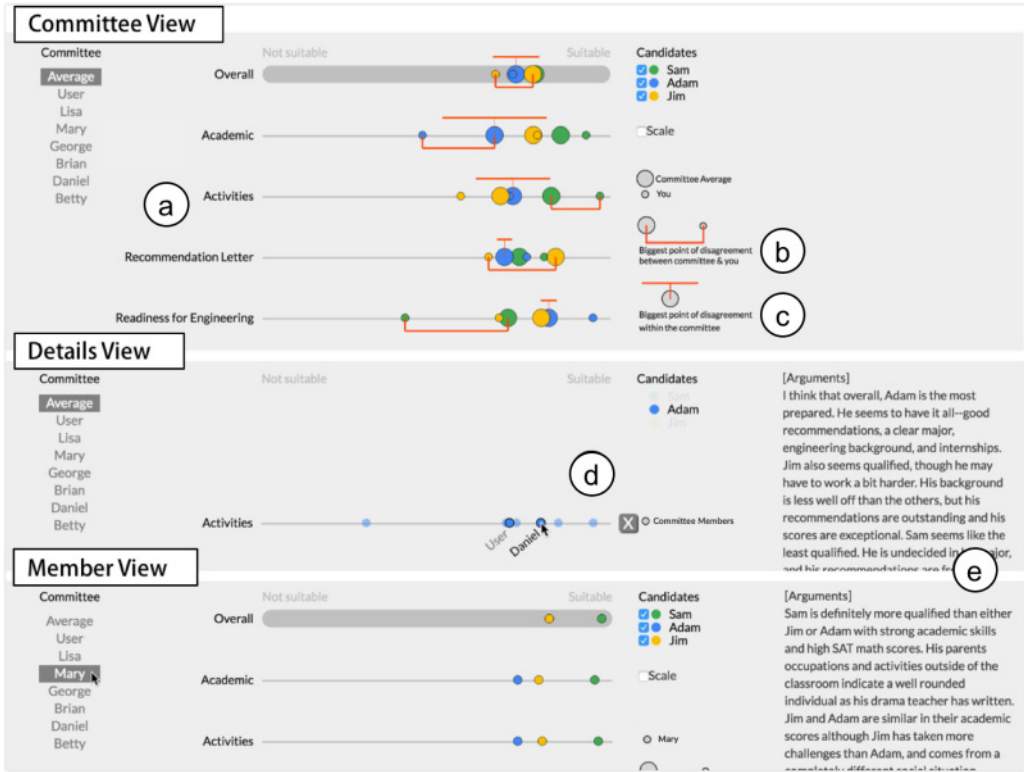
Fig. 2. ConsensUs interactive group visualization interface showing different views. The Committee View (top) shows (a) the average ratings for each criteria as well as the overall average rating. The big dots represent the group average while the small dots represent the score given by the current user. (b) The red line below the scale highlights disagreement between the user and the group. (c) The red line above the big dots shows disagreement (variance in ratings) within the group. Only the biggest disagreement for each criterion is highlighted. By clicking on the big dots, users see the Details View (middle) which shows (d) each group member's relative ratings for a particular criterion and candidate, and each group member's arguments. Finally, users can click member names on the left to see the Member View (bottom) to explore a particular member's opinions.

in the later stages of decision making. As with the independent opinion interface, users can click the scale checkbox to see a 0 to 10 scale along each criterion line.

Highlighting Disagreement. ConsensUs is designed to highlight two different types of disagreement: the variance present within the group as a whole (see Figure 2(c)) and the explicit disagreement between the user and the rest of the group (see Figure 2(b)). Both types of disagreement are represented by red lines (respectively, above and below the criteria bar), which highlight the largest points of disagreements for each criteria. In other words, for each criterion and for the overall average, the system highlights the alternative with the largest group variance and the largest distance between the user and the group average.

Exploring for Details. ConsensUs is interactive, allowing users to explore for more details. By clicking on the large group dots, users can explore how different group members rank alternatives relative to each other (see Details View in Figure 2). From this view, users can look at each

member's arguments in conjunction with their relative ratings (see Figure 2(d)). Users can also use the left column to explore specific group members' opinions. As users hover over a group member's name, they see that member's independent opinion interface (Figure 2 Member View), including their ratings and written arguments.

4 EVALUATION

To evaluate ConsensUs, we designed an experiment that investigates the relative benefits of visualizing disagreements as part of a group decision process. In particular, does the visualization help to supplement written arguments and provide advantages over just relying on arguments alone? Arguments take a more narrative approach to persuasion, making it easier to understand the underlying rationale (Klein 2012; Kriplean et al. 2007). While arguments are more naturalistic, they are subjective and can be easily misconstrued (Wittgenstein 2010). Quantifying subjective opinions can allow for more precise and holistic views of complex decisions, but it does not explicitly encode the rationale or assumptions behind an opinion.

Prior work that compares different representations of data provides some indication of the tradeoffs. For instance, a meta-analysis by Allen and Preiss (1997) compared the persuasiveness of narrative arguments and quantitative evidence within 15 studies, finding quantitative evidence to be more persuasive overall. However, quantitative data in the context of subjective opinions may create interpretation problems, as group members may use quantitative scales differently (e.g., one person's 4 rating may be interpreted the same as another's 7) (Clark and Watson 1995). Besides, researchers recommend combining both narrative arguments and quantitative information. Allen et al. also found that when a message contained both types of information, it was more persuasive than narrative arguments or quantitative evidence on its own (Allen et al. 2000).

Given this background, we posit the following hypotheses:

H1: Participants can more accurately identify points of disagreement with a visualization of quantified subjective opinions than with written arguments. Our study treats this as a manipulation check and measures this with an objective test and a subjective questionnaire.

H2: Participants will give more concrete rationale to support their decision rationale when viewing written arguments. Our study operationalizes this by analyzing the relative amount of decision reasoning (as opposed to process comments) in participants' response to an open-ended question about their decision.

H3: Participants will align their opinion more with the group opinion when they view both a visualization and written arguments of subjective opinions. Our study will measure this by counting changes in the ranking of alternatives, calculating an overall and relative change in ratings, and by open-coding participants' reasons for changing their ratings.

4.1 Conditions

To evaluate these hypotheses, we conducted a between-subject experiment with 87 participants. We compared three conditions: (a) *Arguments-only* condition: Participants can only see the written arguments from other group members; they do not see scores provided by others (see Figure 3). (b) *Visualization-only* condition: The interface shown to the participants is Figure 2 excluding section (e). (c) *Both* condition: Participants simultaneously see the written arguments and the visualization (as in Figure 2).

4.2 Participants

We recruited 90 participants from Amazon Mechanical Turk. We restricted participation to only people in the U.S. and those who did not participate in our pilot studies. To ensure quality, we
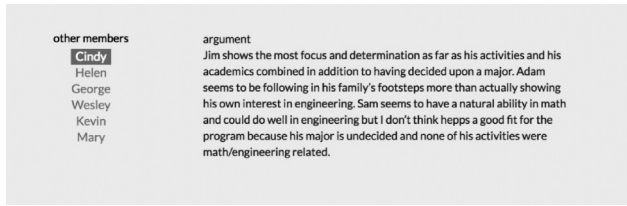
Fig. 3. The group interface for the *Arguments-only* condition.

also restricted recruitment to workers with a 90% HIT (Human Intelligence Tasks) acceptance rate. Each participant received a base compensation of $4.00 for a 30-minute task. We excluded 3 participants due to data incompleteness, leaving 87 participants (33 females, 54 males; age $\mu = 34$, range = 20–74) for our analysis.

4.3 Experimental Task

Participants completed our task individually online using an experimental version of ConsensUs. The experiment was conducted as an individual task rather than a group task for several reasons. We wanted to focus initially on how the visualization affects individual thinking before moving on to groups, which introduce a host of social factors and noise into experimental procedures. Groups are also difficult to recruit and motivate to take decision tasks seriously. This study's focus on how individuals think through the tool can usefully inform how groups will interact.

Adapting materials from Dimicco (2005), we told participants to imagine they were the chair of an admissions committee for an engineering school and the committee must choose one of three candidates to admit into an engineering program. We chose the admissions committee context because the task was: (1) representative of group decisions, (2) open-ended (no right or wrong solutions), (3) complex enough to involve multiple criteria, and (4) accessible enough to run as an online experiment.

Participants received the three candidates' application materials, presented in a spreadsheet with 17 items of information for each candidate (See the Appendix for details). The three candidates were given names with the same gender to avoid gender effects. We also randomly showed the candidates in different orders to counterbalance for order effects. In a pre-survey, only 2 out of 87 indicated that they had previously served on an admissions committee; 40 out of 87 had previously made complex multi-criteria decisions. We explicitly recruited participants who were capable of performing complex reasoning, but did not have prior task experience that could bring about strong opinions on the process or particular criterion. As a post-hoc analysis, we found no significant differences in any outcome measures based on prior experience.

4.4 Confederate Pool

To create a mock decision committee for each participant for our actual study, we recruited 50 participants through Amazon Mechanical Turk beforehand to form a pool of confederates. After accepting the consent form, participants first read the application material. Then, using ConsensUs's independent opinion interface, participants would provide ratings and arguments justifying their relative ratings. Participants were required to write at least 60 words. From these, we selected subjective ratings and arguments from 30 participants. We narrowed down our confederate pool by ensuring: (1) an equal split of preferences and gender for each of the three candidates (i.e., five male and five female confederates favor each candidate), (2) written arguments are reasonable and clear, (3) participants' ratings are commensurate with their written arguments, and (4) each confederate has a clear preference (instead of giving multiple candidates the same ratings).
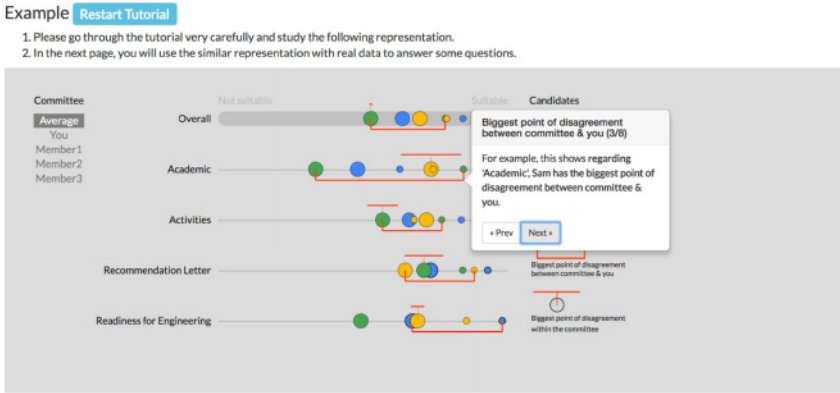
Fig. 4. To train people in the visualization conditions, a step-by-step tutorial guided participants through each feature of ConsensUs.

For each participant in the main experiment, we randomly sampled the confederate pool for an admissions committee of six members, with an equal number of males and females. This led to six-member committees that varied randomly in terms of the number of members who supported each candidate and in terms of how much the committee's average ratings differed from the participants' ratings.

4.5 Procedure

For the main study, after accepting the consent form, participants opened a link that directed them to our experimental platform. The experiment procedure included three phases.

4.5.1 First Phase: Independently Evaluate Three Candidates. All participants were exposed to the independent opinion interface. They were given a walk-through tutorial regarding how to use our visualization. To keep the task consistent across all participants, we pre-selected four criteria (academics, activities, recommendation letters, and fit for engineering) and loaded these in the interface. After providing ratings on each criteria, participants reviewed their overall selection and then filled out two survey questions on a 7-point Likert scale (how confident they are with their opinions and how willing they are to shift their opinions).

4.5.2 Second Phase: View Committee's Opinions. In the group visualization interface, participants viewed the committee's opinion in one of three conditions: *Arguments-only*, *Visualization-only*, or *Both*. First, participants went through a short tutorial with fake data to get familiar with the interface (see Figure 4). Then, participants studied the data from the confederate committee sampled using the method described before. While viewing the committee's opinion, the participants answered eight right or wrong questions about points of disagreements within the committee. Finally, participants answered five Likert-scale questions about the perceived difficulty in identifying points of disagreement.

4.5.3 Third Phase: Provide a Final Independent Evaluation. Participants saw the independent opinion interface again, this time underneath the group interface. They were shown their previous ratings and they were able to change them if they had updated opinions. A series of open-ended questions asked participants to provide reasons for why they changed or did not change their scores, and to provide their opinion about the committee. Participants, again, self-assessed their confidence and willingness to change ratings. Finally, they completed a post-survey that

Table 1. Eight Multiple-Choice Questions in Disagreement Identification Test

Multiple-Choice Questions
Which candidate does the committee support most overall?
Select the candidate that appears to be creating the biggest point of disagreement overall between the committee and you?
Select the criteria that appears to be creating the biggest point of disagreement between the committee and you?
Select the candidate that appears to be creating the biggest point of disagreement within the committee?
Select the criteria that appears to be creating the biggest point of disagreement within the committee?
Overall, which two people on the committee appear to have the most disagreement regarding Sam?
Regarding the "activities" criterion, who on the committee appears to disagree with you most regarding Adam?
Regarding the "readiness for engineering" criterion, who on the committee appears to agree with you most on Jim?

collected their demographics and prior experiences with multi-criteria decisions and admissions committees.

4.6 Measures

Corresponding with our three hypotheses, we measured (1) participants' actual and perceived ability to identify disagreements and the time to complete the task, (2) the relative amount of concrete rationale presented when giving reasons for their decisions, and (3) participants' change in ranking and ratings after viewing the group opinions as well as an open-coding of the decision.

4.6.1 Disagreement Identification Test. To measure H1, participants' ability to identify points of disagreement, we asked eight multiple-choice questions (see Table 1). These questions all had objective answers that could be calculated based on ratings from the participants and the confederate committee. These objective questions served as a manipulation check to ensure that the visualization has the desired effect of helping participants see points of disagreement. We use the sum of correct answers as the participant's score on the disagreement identification test.

As a self-assessment of their ability to reason about the committee, participants were also asked to rate on a 7-point Likert scale (from 1 = Very Difficult to 7 = Very Easy) their ability to: (1) identify points of disagreement within the committee, (2) identify points of disagreement between the committee and themselves, (3) identify committee members with similar opinions, (4) identify committee members with outlier perspectives, and (5) summarize the group opinion.

4.6.2 Decision Rationale. To measure H2, we asked participants to articulate the reasons for their decisions. We phrased the question as *explain why you changed (or did not change) your score* and also asked them to *explain how the committee opinions affect you*. We analyzed this by dividing their written responses into two comment categories: comments on concrete rationale for their decision and comments that discussed their decision-making process/strategy. We operationalized a measure for the amount of decision rationale by calculating the word count and dividing this by

the total number of words in the participant's responses (percent of response devoted to providing rationale). We give an example in the results section below.

4.6.3 Changes in Opinions. To examine H3, we have four measures for changes in opinion. By comparing the first set of independent ratings (before viewing the group opinions) and the second set (after viewing the group opinions), we can calculate the percentage of participants who changed their ratings. Among those who changed ratings, we can calculate the percentage of participants who changed the ranking of candidates (i.e., select a different candidate after viewing the committee opinions). We also calculated the absolute distance between the two sets of ratings (before and after group stage). The absolute ratings change does not account for whether the participant moved closer or farther from the committee. For this, we calculated the rating changes relative to the group average.

To calculate the latter two measures, we represent each set of ratings by a 12-dimensional vector (3 candidates × 4 criteria, corresponding to the dots in the visualization). More formally, we denote the first set of ratings as $p = (p_1, p_2, \ldots, p_{12})$, the second set of ratings as $q = (q_1, q_2, \ldots, q_{12})$, and the set of group average ratings as $g = (g_1, g_2, \ldots, g_{12})$.

— Absolute rating changes. This variable indicates how much participants move their scores from the original scores in the first phase. It is calculated using the Taxicab distance (Krause 2012) of the two 12-dimensional vectors, which can be denoted as $d_{sc}(p, q) = \sum_{i=1}^{12} |p_i - q_i|$.
— Change in ratings toward the committee average. This variable is calculated by the change of distance between participant and committee from before to after viewing group opinions. More formally, $d_g = d_{g\_before}(p, g) - d_{g\_after}(q, g) = \sum_{i=1}^{12} |p_i - g_i| - \sum_{i=1}^{12} |q_i - g_i|$. In this way, if d_g is positive, it means that the distance before is larger than the distance after, which indicates that the participant has shifted their ratings toward the group.

As a covariate in our analysis below, we control for the initial distance between participants' ratings and the committee ratings (start distance from committee). Finally, to give us better qualitative insight for the changes in opinions, we also coded the reasons for their decisions mentioned before. The coding scheme and analysis procedure for this measure is described in the results section below.

5 RESULTS

Participants were fairly distributed across the three candidates as the best choice (37.9%, 34.5%, and 23.0%, respectively, plus 4.6% participants who rated multiple candidates as the best). For the eight disagreement test questions, participants correctly answered 4.01 questions on average ($\sigma = 2.26$). Overall, 36.8% participants changed their ratings to align with the group. Only 5.7% changed their scores to be farther away from the group average. 57.5% did not change their scores to be farther or closer from the group. 16.1% of participants changed their ranking for a different best candidate.

For participants' self-assessment of confidence and willingness to change ratings before and after viewing group opinions, we found that participants had significantly higher confidence after viewing group opinions (one-tailed paired T-test, $t(86) = -1.97, p < 0.05$) and significantly lower willingness to change ratings after viewing group opinions (one-tailed paired T-test, $t(86) = 4.87, p < 0.05$). We found no significant difference across conditions.

5.1 Identifying Disagreements

5.1.1 Visual Highlights Helped Participants Identify Disagreements. We tallied each participant's score on the disagreement identification test. A Kruskal-Wallis test (nonparametric equivalent of one-way ANOVA) shows that there is a significant difference for the number of correct
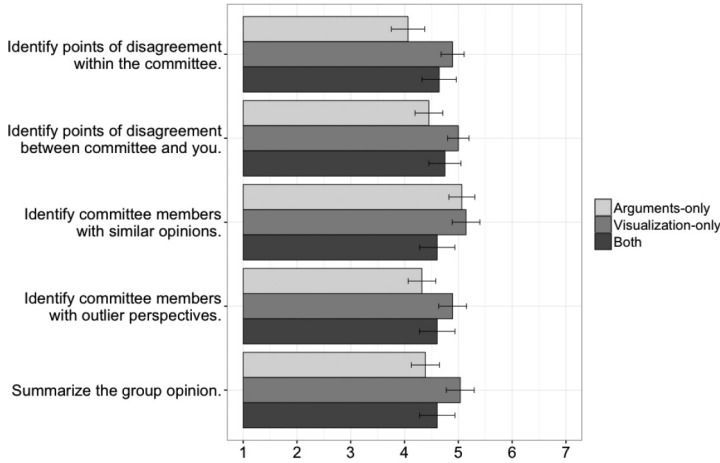
Fig. 5. Subjective evaluation of the ability to identify disagreements and obtain a holistic view of group opinions. Likert scale (1=Extremely Difficult, 7=Extremely Easy).

answers on the eight disagreement identification questions across three conditions, confirming our manipulation check ($\chi^2(2, 87) = 28.13, p < 0.05$) (see Figure 6). Post-hoc Mann-Whitney U pairwise comparisons show that participants in the *Visualization-only* condition ($\mu = 5.00$, $\sigma = 2.05$) and the *Both* condition ($\mu = 4.96$, $\sigma = 2.30$) have a significantly higher correctness rate than the *Arguments-only* condition ($\mu = 2.26$, $\sigma = 1.09$), respectively ($W = 124, p < 0.05; W = 153, p < 0.05$). Chi-Square tests for each question show significantly lower scores for the *Arguments-only* condition, except for the first question, indicating that participants in all conditions were able to identify the committee's favored candidate.

5.1.2 No Significant Difference in Task Complexity. We measured how much time participants needed to perform the Disagreement Identification Test, and ran an ANOVA with the condition as the independent variable. There was no significant effect between conditions (average time in minutes and standard deviation for three conditions: $\mu = 6.78, \sigma = 3.62$ for *Arguments-only*; $\mu = 5.72, \sigma = 2.83$ for *Visualization-only*; $\mu = 6.74, \sigma = 4.83$ for *Both*).

5.1.3 Visualization Participants Found It Easier to Identify Disagreements. After the disagreement identification test, participants answered five self-assessment questions on their ability to reason about the committee and to identify disagreements. Participants in the *Arguments-only* condition generally felt it was more difficult to identify points of disagreement (see Figure 5), again confirming that the visualization had the desired effect of highlighting disagreements. An ordinal logistic regression for the sum of Likert scores of five questions shows a marginal effect of conditions ($\chi^2(2, 87) = 5.929, p = 0.052$). Post-hoc comparisons show that the scores for the *Visualization-only* condition is significantly higher than the *Arguments-only* condition ($Z = 2.413, p < 0.05$), with the *Both* condition falling between the others.

5.2 Arguments Led Participants to Elaborate Their Rationale

After the group stage, participants were asked an open-ended question to articulate the reasons for their decision. We did a grounded, qualitative analysis of this data. We analyzed the percentage of words devoted to concrete rationale versus statements about their process or strategy for making a decision. For example, this participant's first sentence provides process thinking while the second
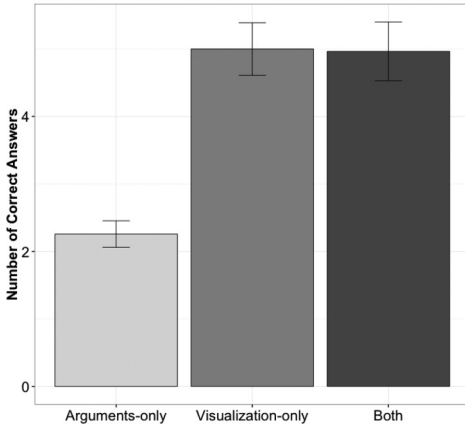
Fig. 6. Participants in the *Visualization-only* and *Both* condition scored higher on the Disagreement Identification Test than the *Arguments-only* condition, confirming our manipulation.
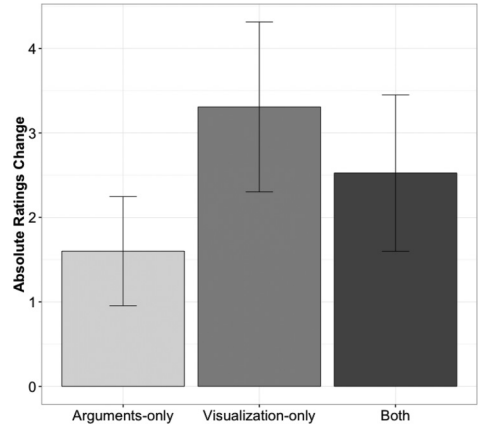
Fig. 7. Absolute Ratings Change. *Visualization-only* changed their scores significantly more than the *Arguments-only* condition.

sentence is counted as concrete rationale: *"I stand by my original rankings based on the information given. All students were capable of entering an engineering program based on their GPA and SAT math scores."* Two members of the research team coded 10% of the participants' arguments while blind to the condition. Agreement between the two coders was computed as the number of matching percentage divided by the total number of sample data. The coders reached an acceptable level of 90% agreement, and a single researcher finished coding the remaining data.

The results show that an average of 51% of the words in each participant's rationale in the *Arguments-only* condition can be categorized as concrete rationale, compared to 28% in *Visualization-only* and 29% in the *Both* condition. In an ANOVA with the condition as an independent variable and the percentage devoted to concrete rationale as a dependent measure finds that there is a significant effect of conditions ($F(2, 85) = 3.337, p < 0.05$).

5.3 Changes in Opinions

5.3.1 Changes in Rank More Likely with Arguments Only. We examined the percentage of participants who changed their ratings and the percentage of participants who changed their ranks. For rating changes, the *Both* condition has the highest percentage (53.6%), and *Visualization-only* has a slightly lower percentage of 46.4%. The *Arguments-only* condition (29.0%) led less people to change their ratings compared to the two conditions with visualization. We saw a different trend in the percentage of participants who actually changed the rank order of their candidates. The two conditions with arguments have similar percentages, 16.1% for *Arguments-only* and 21.4% for *Both*.

5.3.2 Changes in Ratings Less Likely with Arguments Only. We measured how participants changed their ratings from before to after viewing the group opinions in two ways: absolute ratings change and change in ratings toward the committee average. First we looked at the overall change in participants' ratings. An ANOVA, with the condition as an independent variable and start distance from committee as a covariate and the absolute score change as dependent measure, finds a significant effect of the condition ($F(3, 84) = 3.190, p < 0.05$) (see Figure 7). A post hoc Tukey test finds a significant difference between the *Visualization-only* condition ($\mu = 3.31$,

$\sigma = 5.32$) and the *Arguments-only* condition ($\mu = 1.60$, $\sigma = 3.60$) ($p < 0.05$); the *Both* condition was not significantly different from the other two groups, lying somewhere in the middle.

We also examined the distance between participants' ratings and committee ratings from before to after viewing group opinions. The *Both* condition has the highest percentage of people (16.1%) who changed their ratings to align with the group (8.0% for *Arguments-only* and 12.6% for *Visualization-only*). An ANOVA with the condition as an independent variable and start distance from the committee as a covariate, and the change in ratings toward the committee average as a dependent measure finds a significant effect of condition ($F(3, 84) = 4.106, p < 0.05$). The *Visualization-only* and the *Both* condition saw significant change in ratings toward the group average ($p = 0.022$ and $p = 0.027$, respectively). Participants in the *Arguments-only* condition did not significantly change their ratings.

5.3.3 Reasons for Changing Ratings after Viewing Committee Opinions. In addition to coding the percentage of concrete rationale mentioned above, we also identified eight key reasons (three reasons why their ratings changed and five for why they did not change) from participants' decision rationale. Two members of the research team coded 10% of the data while blind to condition. Agreement between the two coders was computed as the number of matching categorizations divided by the total number of possible categorizations. Thus, an agreement was recorded only when an assigned category matched for two coders. The coders reached an acceptable level of 90% agreement, and a single researcher finished coding the remaining data.

Figure 8 presents the comment-coding scheme with examples and a percent breakdown of reasons across conditions. The percentage is the number of people who express the particular reason divided by the number of participants in each condition. Some arguments are categorized to multiple reasons. A couple of trends emerge from these data that are further exemplified by the qualitative data below. First, we explore reasons for why participants seemed to change their opinions followed by why participants did not change their perspective.

Participants across Conditions Were Influenced by Information and Opinions Provided by the Group. Roughly equal numbers of participants across conditions talked about changing their score because they learned new information (7%) or because the committee's opinions changed their perspective (26%). For example:

> I took what the other members said and realized that it would probably be better to choose a candidate that was more committed and focused on engineering. (*Arguments-only*)

> After seeing the committee results, I realized that I had probably been a bit harsh on Sam due to the alcohol offense. (*Visualization-only*)

> I changed my score a bit to reflect more of what the committee thought. I thought that they had some good points, so I changed a couple of things on my scores. (*Both*)

From these categories of reasons, we did not see differences between conditions in terms of participants' willingness to change their ratings based on new insights that arose during the group stage.

Visualizing Disagreements Led Participants to Change Ratings but Not Their Rankings. Participants expressed another reason for changing their ratings: they wanted their ratings to be closer to the committee, even though their original assessment did not change. This notion of a "ratings hedge" to align with the committee without changing the overall order of their decision happened more in the *Visualization-only* (14%) and *Both* conditions (18%), than the *Arguments-only* condition (6%).

Changed Ratings

Reasons	Quote	A	V	B	Total
Affected by the group opinion	*After seeing the committee results, I realized that I had probably been a bit harsh on Sam due to the alcohol offense.*	26%	21%	32%	26%
Notice new information	*The points they brought up about Adam's lack of challenging courses in academics made me realize I had scored him to high in that area.*	6%	7%	7%	7%
Changed but still hold original opinion	*The committee results really did not change my original assessment, I only adjusted on the scale accordingly.*	6%	14%	18%	13%

Did Not Change Ratings

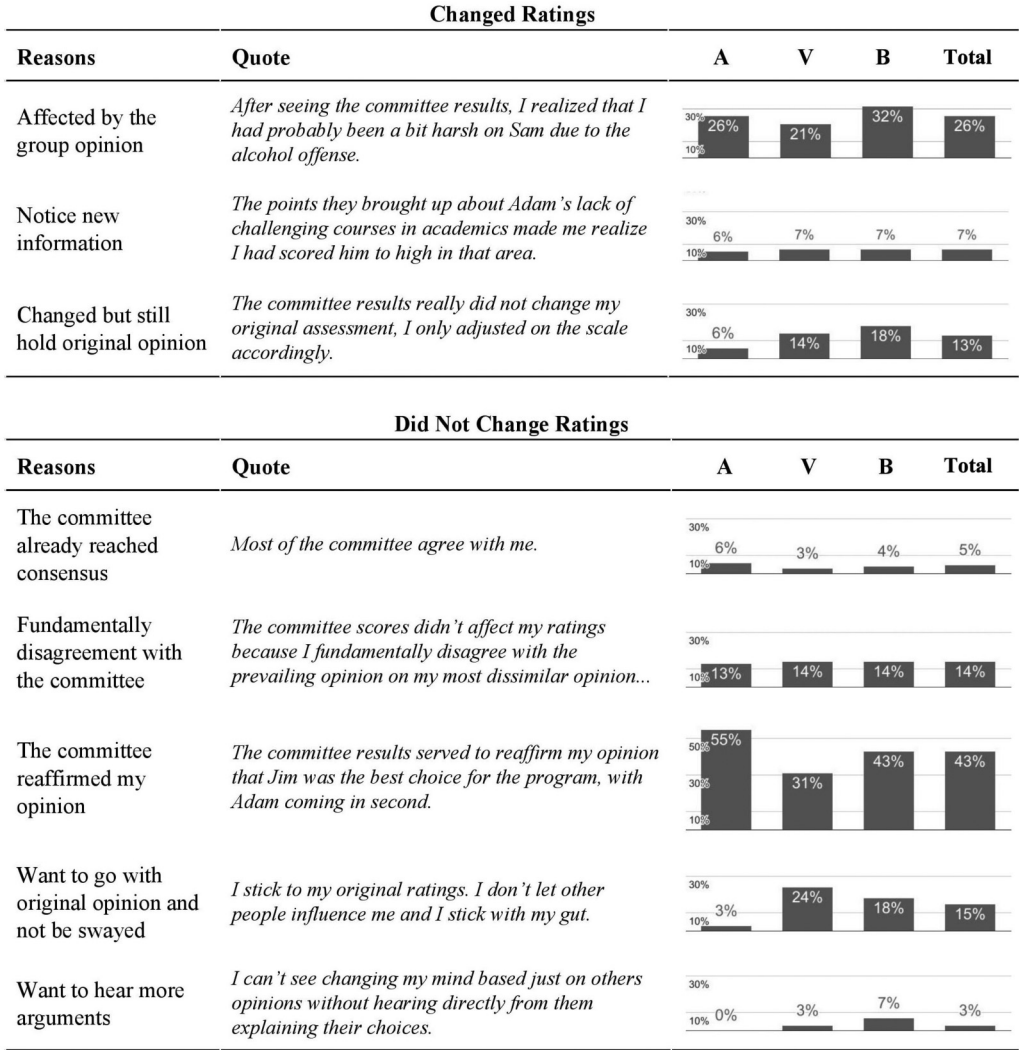| Reasons	Quote	A	V	B	Total
The committee already reached consensus	*Most of the committee agree with me.*	6%	3%	4%	5%
Fundamentally disagreement with the committee	*The committee scores didn't affect my ratings because I fundamentally disagree with the prevailing opinion on my most dissimilar opinion…*	13%	14%	14%	14%
The committee reaffirmed my opinion	*The committee results served to reaffirm my opinion that Jim was the best choice for the program, with Adam coming in second.*	55%	31%	43%	43%
Want to go with original opinion and not be swayed	*I stick to my original ratings. I don't let other people influence me and I stick with my gut.*	3%	24%	18%	15%
Want to hear more arguments	*I can't see changing my mind based just on others opinions without hearing directly from them explaining their choices.*	0%	3%	7%	3%

Fig. 8. Eight key categories of reasons participants provided for changing or not changing their ratings, with example quotes, and percentages for each condition and overall. A: *Arguments-only*, V: *Visualization-only*, B: *Both*, T: *Total*. One participant's written arguments may be categorized to multiple reasons.

> I felt I must have missed some of the information if my scores were off a lot so I changed them a little to be more like the other participants. I was pretty happy with my scores originally. (*Visualization-only*)

> I changed my score to match the scoring areas of the others. As far as my beliefs in the order in which the three finished in my opinion, that stayed the same. (*Both*)

As the quotes indicate, the precise visualization of relative scores afforded participants this ability to better align their ratings even if they stick to their original opinion.

Participants Did Not Change Ratings with High Initial Agreement or Disagreement. Looking now at reasons why participants did not change, it was clear that the amount of initial alignment affected

the ratings. If the participant's ratings already aligned with the committee, then participants tended not to change their score and reported they had already reached consensus, and this occurred in similar amounts across conditions. On average, 5% of participants in every condition mentioned this:

> Most of the committee agrees with me. (*Arguments-only*)

> If there were huge discrepancies I would have paid more attention to them, but it was fairly close. (*Visualization-only*)

> I originally chose the candidate based on their best attributes, which also seemed to agree with the results of the committee. (*Both*)

On the other hand, for initially wide disagreements between the group and the participant, some people just reported that they fundamentally disagreed with the others. Roughly 13% of people in each condition said they disagreed without giving a specific reason:

> The committee raised thoughts worth considering, but I can't say they had significant impact on my overall view of each, or as a group. Re-reading the info given on them, I can't say it changes my mind. (*Arguments-only*)

> I didn't make any changes. The committee scores didn't affect my ratings because I fundamentally disagree with the prevailing opinion. (*Visualization-only*)

> I understand and respect their decisions and opinions, but they do not concur with my own. (*Both*)

For these participants, the large gap between their opinion and the group was too much to sway their ratings. While condition did not affect these participants, the presence or absence of arguments did lead other participants to reveal reasons for not changing their ratings that help us understand the interface effects.

Arguments-Only Reaffirmed Many Participants' Initial Beliefs. On average about 43% of participants commented they did not change their score specifically because of the committee's opinions: the ratings and reasoning from others seemed to reaffirm their initial beliefs. The general trend shows that more people in the *Arguments-only* condition (55%) compared to the *Visualization-only* condition (31%) expressed this notion of reaffirming their beliefs. The *Both* condition (43%) fell somewhere between. Many participants in the *Arguments-only* condition expressed that they felt more confident in their initial analysis:

> I made my decision based on the information presented on the spreadsheet... that information is just as, if not more, pertinent than the views of the committee members. (*Arguments-only*)

Those participants tended to refer back to the candidate information for why they did not change their opinion. Other participants in the *Arguments-only* condition seem to seek confirmatory evidence in others' arguments, possibly a sign of confirmation bias (Wason 1960).

> The committee results served to reaffirm my opinion that Jim was the best choice... The committee's comments showed me that I was most likely correct in my opinion. (*Arguments-only*)

These participants seemed to explicitly seek out other committee members and information that supported their opinion and ignored opinions that did not align with their initial judgments.

Visualization Led Participants to be More Aware of Disagreements. In general, participants who viewed the visualization were more cognizant of precise differences between their ratings and those of the committee. The visualization increased their performance on the Disagreement Identification Test and their perceived ability to find disagreements. It was also evident in their comments:

> Upon reviewing my scores against the committee, I could review where the discrepancies were and reevaluate my own scores. (*Visualization-only*)

In contrast, participants who only saw arguments discussed how difficult it was to identify the crux of committee disagreements:

> It was extremely difficult because most of the committee members are essentially saying the same things in different ways. It was hard to pinpoint a clear source of disagreement within the committee because of this. (*Arguments-only*)

Interestingly, an increased awareness led some visualization participants to report that they did not want their opinion to be swayed by the quantitative differences. From Table 1, we see 24% of *Visualization-only* and 18% of *Both* participants articulated that they did not want to be blindly swayed by other committee members, compared to only 3% in the *Arguments-only* condition.

> I chose those before I knew anyone else's opinion. Now that I know what others rated these individuals, it would sway my opinion one way or another. (*Visualization-only*)

> I stuck to my original ratings. I didn't let other people influence me and I stuck with my gut. (*Visualization-only*)

So this "ratings awareness" led some participants to resist changing their ratings. Others, as described above, were more willing to make rating changes to align with the committee or to appear to align better with the committee. Next, we reflect on whether the presence of such a visualization actually influences the consensus building process.

6 DISCUSSION

Our evaluation of ConsensUs explored the relative effects of quantifying and visualizing subjective opinions. Specifically, we investigated different representations of opinions during a multi-criteria group decision task and investigated how participant's ability to identify disagreements affected the decision rationale and amount of change in opinions after viewing the committee's opinion. We revisit the hypotheses, interpret the findings within the group decision-making literature, and outline some future work.

6.1 Hypotheses Revisited

The results support H1 and H2, with mixed support for H3.

6.1.1 H1: Visualization Helps Participants Identify Points of Disagreement. As a confirmation of our main independent manipulation, we found that when provided a visualization of quantified subjective opinions (in the *Visualization-only* and *Both* conditions), participants could more accurately identify points of disagreement than written arguments. Participants also self-reported feeling they could identify points of disagreement easier in both the *Visualization-only* and *Both* conditions than *Arguments-only* condition. Also, in open-ended comments, participants with the visualization said they could identify discrepancies within the group better, while participants who just viewed arguments struggled to pinpoint the disagreements.

6.1.2 H2: Arguments Support Concrete Reasoning. Supporting H2, we found that participants in the *Arguments-only* condition wrote rationale that emphasized concrete evidence, and focused less on their process of reaching a decision. Participants in the *Visualization-only* and *Both* conditions seemed to focus more on the mechanics of how they made their decisions in relation to the group. Interestingly, from the qualitative analysis, we found that participants in the *Arguments-only* condition tended to re-affirm their beliefs rather than consider alternatives more than the other conditions. This might be an indication that participants more often stay entrenched in their original views when viewing arguments alone; the disagreement awareness afforded by the visualization could be helping to elide anchoring effects (Tversky and Kahneman 1975).

Additionally, the result that participants wrote more elaborate rationale in *Arguments-only* but not in the *Both* condition implies that arguments did not affect participants' reasoning as much when presented together with the visualization. We examined the process-related rationale and found that the visualization dominated most of the *Both* participants' reasoning (of 28 participants, 18 referenced the visualization; 3 only referenced the arguments; 7 gave ambiguous arguments). Further studies can investigate how to connect the visualization and arguments more productively.

6.1.3 H3: Visualization and Arguments Better for Aligning to Group Opinion. Given the ability to more clearly see points of disagreement, we found that participants in the visualization conditions (*Visualization-only* and *Both*) were more likely to change their ratings toward the committee. This supports prior work by Osatuyi et al. (2015) where making preferences visible led participants to change their preferences toward the group. However, the qualitative analysis revealed that people changed their scores for a variety of reasons. Some participants genuinely changed their score to align more with the group (same across conditions). Some tended to hedge or normalize their ratings to align more with the group, even though they did not actually change their opinion; this normalization occurred more with the visualization than in the *Arguments-only* condition. This conformity could be interpreted positively as crafting consensus or could be a sign of groupthink where group members accept the prevailing opinion without thoughtful debate (Janis 1972).

Interestingly, participants in the *Arguments-only* condition changed their rankings more, but also reported to confirm their previously held opinions more than the other conditions. This reaffirmation of existing beliefs could lead argumentation-only groups to confirmation bias (Wason 1960; Nickerson 1998; Toma and Butera 2009). It could be that *Arguments-only* creates a more bimodal interaction where participants either dramatically changed their view or stick to previously held beliefs), while the visualization supports a more nuanced representation of one's opinion.

6.2 Do Participants Reach Consensus?

ConsensUs seems, at some level, to help support early stages of consensus building as outlined by Briggs et al. (2005): participants can more easily diagnose the disagreements and then read arguments from others that articulate the reasons for those disagreements. The visualization led to more changes in ratings toward the group, but participants who only saw arguments changed rankings more. The results show that quantified opinions and written arguments provide a different value for understanding the group, but our data do not untangle whether these rating changes and the reasons provided constitute consensus building, or some other phenomenon like false consensus (Mullen et al. 1985).

On one hand, the visualization helped expose the disagreement in the group and potentially led people to be less anchored to their initial opinions. This finding is supported in Figure 8 where we see more participants in the *Arguments-only* condition felt that the committee comments only reaffirmed their opinions. Similarly, more participants in the *Arguments-only* condition believed the committee already reached consensus and essentially ignored the different opinions in their

group. In the other conditions, the visualization helped participants identify points of disagreement by essentially lowering the cognitive load required to understand where the disagreements lie.

On the other hand, our data also indicate that the visualization often led participants to appease the group. We saw a high percentage of visualization participants who *changed their score but still held to their original opinion.* Such appeasement can be problematic if it prevents independent thinking. However, we argue that the presence of technology does not inherently make people more likely to appease the group (this phenomenon certainly also happens in non-technology-mediated settings), but rather frees up resources to grapple with the disagreements. As discussed by Briggs et al. (2005), reaching consensus still requires thoughtful debate and measured compromise. One could imagine how a tool like ConsensUs could be extended to support subsequent steps in the consensus building process.

Our study provides initial insights about how quantified opinions and written arguments could affect people's decisions and offers lessons from our visualization design. Future studies will focus on mitigating the potential negative effects and examining decisions in an actual group setting to better understand whether such an intervention helps groups reach consensus more effectively and with higher satisfaction.

6.3 Limitations

Like many lab experiments, this study does not perfectly replicate the real world (McGrath 1984). The study was performed online with Mechanical Turk workers on a simulated task, rather than in an actual group-decision setting. While Mechanical Turk workers typically do not serve on university admissions, we found their skills and comprehension to be sufficient enough for this study. The simulated scenario still provided useful insights into how visual aids can focus decision-makers' attention on the crux of disagreements.

Further, the online setting minimized social factors—such as status (Dubrovsky et al. 1991) and social loafing (Latane et al. 1979)—that typically arise in face-to-face group decisions. Participants were likely less invested in the decision outcome than real-world decision scenarios. This lack of investment may have led participants, for example, to normalize their ratings toward the group more than they would have for real decisions. This study contributes an initial evaluation of the tool itself before gathering evidence in the wild.

While our research team did not tell people we were measuring pre- and post-group alignment, some participants may have inferred this from a system that visualizes disagreement. This may have created demand characteristics that influenced visualization participants to move more toward the group. That said, participants were told to shift their opinion however they wanted, and indeed, many stuck to their original score or even shifted farther away from the group opinion.

6.4 Future Work

6.4.1 Extend Disagreement Identification. The results demonstrate the value of utilizing visualization to externalize and identify points of disagreement. Participants in the *Arguments-only* condition expressed frustration when trying to identify disagreement, while the visualization provided an intuitive way to explore differences in the group. As one participant in the Both condition said, "The interactive chart was very user friendly and allowed me to look at individual scoring and compare to the group scoring. It made locating disagreements easy."

In addition to identifying disagreement through quantification of subjective opinions, one could potentially identify conflict through automated text-analysis. Work on topic modeling in arguments provides a strategy for augmenting deliberation technologies (Chuang et al. 2012; Klein 2012; Kriplean et al. 2012a, 2012b).

6.4.2 Explore Task Complexity. Admitting prospective students for an engineering program is fairly limited in complexity. Further studies can explore how tasks of different complexities can affect the consensus building process; visualization may provide more value when there are greater number alternatives and criteria to consider. Complexity also increases with the number of group members. With more group members, it may be even harder to identify points of disagreement or obtain a holistic view of group through arguments alone. Large numbers of group members might influence members' tendency to fall into groupthink or remain rooted in their own views. Researchers have also found that large diverse groups make better judgments in aggregate (Dalal et al. 2011). How to design and leverage such a tool at scale is a fruitful area to explore.

6.4.3 Investigate in More Diverse Settings. By explicitly highlighting the points of disagreement, a tool like ConsensUs has potential to help groups effectively discover information that might be hidden or miscommunicated to some group members (Stasser and Titus 2003). Conflict highlighting could also be useful for facilitating collective sensemaking (Goyal and Fussell 2016).

The visualization approach embodied by ConsensUs also has potential to support asynchronous communication for distributed groups. Asynchronous communication allows groups to discuss wider range of topics in a parallel fashion and pool more information together (Dalal et al. 2011). Combining both synchronous and asynchronous communication may allow group members to devote more effort to their decision and thus achieve better outcomes. Relevant techniques from discussion forums can be adapted into the tool.

As mentioned earlier, future studies should engage real-world groups to more deeply understand the value of a tool like ConsensUs. Real-world decision-making processes will be more complicated and thus have more factors that will affect the results. For example, rather than having a default set of criteria, real-world groups need to first build consensus on potential criteria before considering candidates. How groups delegate, weight, and choose different criteria can also be important for the decision results.

7 CONCLUSION

We introduced ConsensUs, a real-time interactive visualization tool for supporting groups in the consensus building process. ConsensUs explicitly highlights subjective points of group disagreement across multiple criteria while presenting participants with other group member's arguments. We evaluated ConsensUs in a between-subjects experiment with 87 participants that compared three conditions: written discourse only, visualization only, and written discourse plus visualization. We compared the participants' ability to identify disagreement as part of a mock decision committee and their subsequent changes to their individual opinions. Based on the analysis and study results, we found that visual support helps participants identify points of disagreement and leads to final assessments more aligned with the group. Participants align more for a variety of reasons including both group appeasement and genuine agreement. Future work involves extending disagreement identification to text-based arguments and investigating the tool in more diverse settings.

APPENDIX

This appendix shows the candidates' information given to the experiment participants. (adapted from Dimicco (2005))

<div align="center">Sam Smith</div>

(1) Hometown: Dayton, Ohio
(2) Private high school

(3) SAT VERBAL 790
(4) SAT MATH 770
(5) GPA 3.8 out of 4.0
(6) Sam is taking three Advanced Placement (AP) classes and one non-AP classes.
(7) He is taking three Science/Engineering related AP classes including Calculus, Statistics, and Computer Science.
(8) Previous students from his school who attended UCSD did very well academically and socially.
(9) He plays varsity football and is a member of the debate team.
(10) Sam and his teammates won first place in a High School Innovation Challenge, which asked student teams to design a product.
(11) His father is a CTO (chief technology officer) of a startup and his mother is a consultant at a big tech company.
(12) Sam's sister attends Ohio State University.
(13) He says he is undecided in his major.
(14) A note on his transcript indicates that Sam was suspended from school for 3 days for consuming alcohol during a school field trip.
(15) Sam's football coach says in a recommendation that he is a very talented player and could compete at the college level.
(16) His drama teacher says in his recommendation that Sam has shown a natural talent for acting and he is a very well-rounded student.

Adam Adams

(1) Hometown: Los Angeles, CA
(2) Large high school, high-to-middle income neighborhood
(3) SAT VERBAL 600
(4) SAT MATH 730
(5) GPA 3.5 out of 4.0
(6) He is taking five courses this term.
(7) He has only taken regular (non-AP) courses (including the five courses he's taking this term).
(8) His high school has a high reputation for being a high-quality school with a tough grading policy.
(9) He is one of the leaders of the engineering club in his school.
(10) He won first place in the annual state-wide science fair.
(11) Both of his parents are senior engineers in a big technology company.
(12) His brother attends UCLA.
(13) He intends to major in electrical engineering.
(14) "Adam's physics teacher has written a letter of recommendation saying that he is a proactive student who is able to successfully develop plans and implement them."
(15) He worked two summers at two electrical engineering firms. The manager of one of the companies has written a letter of recommendation for him.
(16) His summer intern manager's letter of recommendation is way too short compared to the common length.

Jim Jones

(1) Hometown: Newton, Massachusetts
(2) Large public high school, middle-to-low income neighborhood.

(3) SAT VERBAL 680

(4) SAT MATH 750

(5) GPA 3.6 out of 4.0

(6) He is taking three AP classes and one non-AP class

(7) He has avoided non-academic courses such as home economics and health because he thinks it's a waste of time.

(8) Previous students from his school who attended UCSD had some academic difficulties.

(9) Jim took classes at his local community college during the summers.

(10) He won first place in his school's math competition.

(11) His parents are divorced. Mother works in retail. Father works in construction.

(12) Jim has an elder brother who only has a high school degree.

(13) He intends to major in math-related majors (e.g., engineering, physics, architecture)

(14) His math teacher has written a letter of recommendation saying that Jim is extremely good at math and he is well prepared for any math-related majors.

(15) His English teacher mentions that he frequently handed his homework in late.

(16) His high school principal has written a letter of recommendation saying that her high school has a strong academic reputation, and describes Jim as a hard-working student.

REFERENCES

Mike Allen, Rebecca Bruflat, Renée Fucilla, Michael Kramer, Steve McKellips, Daniel J. Ryan, and Marieke Spiegelhoff. 2000. Testing the persuasiveness of evidence: Combining narrative and statistical forms. *Communication Research Reports* 17, 4 (2000), 331–336.

Mike Allen and Raymond W. Preiss. 1997. Comparing the persuasiveness of narrative and statistical evidence using meta-analysis. *Communication Research Reports* 14, 2 (1997), 125–131.

Sergio Alonso, Enrique Herrera-Viedma, Francisco Javier Cabrerizo, Francisco Chiclana, and Francisco Herrera. 2007a. Visualizing consensus in group decision making situations. In *2007 IEEE International Fuzzy Systems Conference*. IEEE, 1–6.

Sergio Alonso, Enrique Herrera-Viedma, Francisco Javier Cabrerizo, Carlos Porcel, and Antonio Gabriel López-Herrera. 2007b. Using visualization tools to guide consensus in group decision making. In *International Workshop on Fuzzy Logic and Applications*. Springer, 77–85.

Gennady L. Andrienko and Natalia V. Andrienko. 2002. Interactive visual tools for spatial multicriteria decision making. In *Proceedings of the Working Conference on Advanced Visual Interfaces*. ACM, 129–132.

Bon Adriel Aseniero, Tiffany Wun, David Ledo, Guenther Ruhe, Anthony Tang, and Sheelagh Carpendale. 2015. STRATOS: Using visualization to support decisions in strategic software release planning. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1479–1488.

Michel Avery, B. Auvine, B. Striebel, and L. Weiss. 1981. A handbook for consensus decision making: Building united judgement. Madison, WI: The Center for Conflict Resolution.

Sanjana Bajracharya. 2014. *Interactive Visualization for Group Decision-making*. Ph.D. Dissertation. University of British Columbia.

Aruna D. Balakrishnan, Susan R. Fussell, and Sara Kiesler. 2008. Do visualizations improve synchronous remote collaboration? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1227–1236.

Boris B. Baltes, Marcus W. Dickson, Michael P. Sherman, Cara C. Bauer, and Jacqueline S. LaGanke. 2002. Computer-mediated communication and group decision making: A meta-analysis. *Organizational Behavior and Human Decision Processes* 87, 1 (2002), 156–179.

Raquel Benbunan-Fich, Starr Roxanne Hiltz, and Murray Turoff. 2003. A comparative content analysis of face-to-face vs. asynchronous group decision making. *Decision Support Systems* 34, 4 (2003), 457–469.

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

Giuseppe Carenini and John Loyd. 2004. ValueCharts: Analyzing linear models expressing preferences and evaluations. In *Proceedings of the Working Conference on Advanced Visual Interfaces*. ACM, 150–157.

Chen-Tung Chen. 2000. Extensions of the TOPSIS for group decision-making under fuzzy environment. *Fuzzy Sets and Systems* 114, 1 (2000), 1–9.

Richard Chimera. 1992. Value bars: An information visualization and navigation tool for multi-attribute listings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 293–294.

Jason Chuang, Christopher D. Manning, and Jeffrey Heer. 2012. Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 74–77.

Lee Anna Clark and David Watson. 1995. Constructing validity: Basic issues in objective scale development. *Psychological Assessment* 7, 3 (1995), 309.

Cristina Conati, Giuseppe Carenini, Enamul Hoque, Ben Steichen, and Dereck Toker. 2014. Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. In *Computer Graphics Forum*, Vol. 33. Wiley Online Library, 371–380.

Siddhartha Dalal, Dmitry Khodyakov, Ramesh Srinivasan, Susan Straus, and John Adams. 2011. ExpertLens: A system for eliciting opinions from a large pool of non-collocated experts with diverse knowledge. *Technological Forecasting and Social Change* 78, 8 (2011), 1426–1444.

Norman Crolee Dalkey, Bernice B. Brown, and Samuel Cochran. 1969. *The Delphi Method: An Experimental Study of Group Opinion*. Vol. 3. Rand Corporation, Santa Monica, CA.

Alan R. Dennis, Kelly M. Hilmer, and Nolan J. Taylor. 1997. Information exchange and use in GSS and verbal group decision making: Effects of minority influence. *Journal of Management Information Systems* 14, 3 (1997), 61–88.

Joan Morris Dimicco. 2005. *Changing Small Group Interaction Through Visual Reflections of Social Behavior*. Ph.D. Dissertation. Massachusetts Institute of Technology.

Vitaly J. Dubrovsky, Sara Kiesler, and Beheruz N. Sethna. 1991. The equalization phenomenon: Status effects in computer-mediated and face-to-face decision-making groups. *Human-computer Interaction* 6, 2 (1991), 119–146.

Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. 2010. Opinion space: A scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1175–1184.

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*. ACM, 299–308.

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.

Tomas Gal, Theodor Stewart, and Thomas Hanne. 2013. *Multicriteria Decision Making: Advances in MCDM Models, Algorithms, Theory, and Applications*. Vol. 21. Springer Science & Business Media.

Theodore J. Gordon. 2009. The real-time Delphi method. *Futures Research Methodology Version* 3 (2009), 19.

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*. ACM, 288–302.

Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. 2013. Lineup: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2277–2286.

Tobias Greitemeyer and Stefan Schulz-Hardt. 2003. Preference-consistent evaluation of information in the hidden profile paradigm: Beyond group-level explanations for the dominance of shared information in group decisions. *Journal of Personality and Social Psychology* 84, 2 (2003), 322.

J. Richard Hackman and Robert E. Kaplan. 1974. Interventions into group process: An approach to improving the effectiveness of groups. *Decision Sciences* 5, 3 (1974), 459–480.

William A. Haskins and Andrew H. Nilssen. 2015. The collaborative enterprise. Retrieved from http://cp.wainhouse.com/content/the-collaborative-enterprise-paper.

F. Herrera, E. Herrera-Viedma, and J. L. Verdegay. 1996. A model of consensus in group decision making under linguistic assessments. *Fuzzy Sets and Systems* 78, 1 (1996), 73–87.

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.

Andrea B. Hollingshead. 1996. The rank-order effect in group decision making. *Organizational Behavior and Human Decision Processes* 68, 3 (1996), 181–193.

Hiroko Itakura. 2001. Describing conversational dominance. *Journal of Pragmatics* 33, 12 (2001), 1859–1880.

Irving L. Janis. 1972. Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes. (1972).

Karen A. Jehn. 1997. A qualitative analysis of conflict types and dimensions in organizational groups. *Administrative Science Quarterly* (1997), 530–557.

Paul E. Jones and Peter H. M. P. Roelofsma. 2000. The potential for social contextual and group biases in team decision-making: Biases, conditions and psychological mechanisms. *Ergonomics* 43, 8 (2000), 1129–1152.

Ralph L. Keeney. 1996. Value-focused thinking: Identifying decision opportunities and creating alternatives. *European Journal of Operational Research* 92, 3 (1996), 537–549.

Daniel A. Keim, Ming C. Hao, Umesh Dayal, and Meichun Hsu. 2002. Pixel bar charts: A visualization technique for very large multi-attribute data sets. *Information Visualization* 1, 1 (2002), 20–34.

Sara Kiesler and Lee Sproull. 1992. Group decision making and communication technology. *Organizational Behavior and Human Decision Processes* 52, 1 (1992), 96–123.

Mark Klein. 2012. Enabling large-scale deliberation using attention-mediation metrics. *Computer Supported Cooperative Work (CSCW)* 21, 4–5 (2012), 449–473.

Eugene F. Krause. 2012. *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*. Courier Corporation.

Travis Kriplean, Ivan Beschastnikh, David W. McDonald, and Scott A. Golder. 2007. Community, consensus, coercion, control: cs* w or how policy mediates mass participation. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work*. ACM, 167–176.

Travis Kriplean, Jonathan Morgan, Deen Freelon, Alan Borning, and Lance Bennett. 2012a. Supporting reflective public thought with considerit. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. ACM, 265–274.

Travis Kriplean, Michael Toomim, Jonathan Morgan, Alan Borning, and Andrew Ko. 2012b. Is this what you meant?: Promoting listening on the web with reflect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1559–1568.

Simon S. K. Lam and John Schaubroeck. 2000. Improving group decisions by better pooling information: A comparative advantage of group decision support systems.*Journal of Applied Psychology* 85, 4 (2000), 565.

Helmut Lamm and Gisela Trommsdorff. 1973. Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review. *European Journal of Social Psychology* 3, 4 (1973), 361–388.

Bibb Latane, 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.

Harold A. Linstone and Murray Turoff (Eds.). 1975. *The Delphi Method: Techniques and Applications*. Vol. 29. Addison-Wesley, Reading, MA. https://s3.amazonaws.com/academia.edu.documents/44240005/TheDelphiMethodTofC.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1513968768&Signature=56n4gUmdPpA01CzpJ%2BNfI5XzCH0%3D&response-content-disposition=inline%3B%20filename%3DThe_Delphi_Method_Techniques_and_Applica.pdf.

Loomio.org. 2017. Homepage. Retrieved May 4, 2017 from http://www.loomio.org.

Russell Mannion and Carl Thompson. 2014. Systematic biases in group decision-making: Implications for patient safety. *International Journal for Quality in Health Care* 26, 6 (2014), 606–612.

Gary Marks and Norman Miller. 1987. Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin* 102, 1 (1987), 72.

J. E. McGrath and A. B. Hollingshead. 1990. Effects of technological enhancements on the flow of work in groups: Preliminary report of a systematic review of the research literature (Report 90-1). *Urbana: University of Illinois* (1990).

Joseph Edward McGrath. 1984. *Groups: Interaction and Performance*. Vol. 14. Prentice-Hall Englewood Cliffs, NJ.

Gilberto Montibeller and Detlof Winterfeldt. 2015. Cognitive and motivational biases in decision and risk analysis. *Risk Analysis* 35, 7 (2015), 1230–1251.

Brian Mullen, Jennifer L. Atkins, Debbie S. Champion, Cecelia Edwards, Dana Hardy, John E. Story, and Mary Vanderklok. 1985. The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology* 21, 3 (1985), 262–283.

David G. Myers and Helmut Lamm. 1976. The group polarization phenomenon. *Psychological Bulletin* 83, 4 (1976), 602.

Raymond S. Nickerson. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2, 2 (1998), 175.

Chitu Okoli and Suzanne D. Pawlowski. 2004. The Delphi method as a research tool: An example, design considerations and applications. *Information & Management* 42, 1 (2004), 15–29.

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.

Polleverywhere. 2017. Homepage. Retrieved May 4, 2017 from http://www.polleverywhere.com/.

Patrick Riehmann, Manfred Hanfler, and Bernd Froehlich. 2005. Interactive sankey diagrams. In *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS'05)*. IEEE, 233–240.

Rosa Romero, David Díez, Kent Wittenburg, and Paloma Díaz. 2012. Envisioning grid vulnerabilities: Multi-dimensional visualization for electrical grid planning. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 701–704.

Louis B. Rosenberg. 2015. Human swarms, a real-time paradigm for collective intelligence. *Collective Intelligence* (2015). Santa Clara CA.

Jeffrey B. Schmidt, Mitzi M. Montoya-Weiss, and Anne P. Massey. 2001. New product development decision-making effectiveness: Comparing individuals, face-to-face teams, and virtual teams. *Decision Sciences* 32, 4 (2001), 575–600.

Hong Sheng, Fiona Fui-Hoon Nah, and Keng Siau. 2005. Strategic implications of mobile technology: A case study using value-focused thinking. *Journal of Strategic Information Systems* 14, 3 (2005), 269–290.

Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of IEEE Symposium on Visual Languages*. IEEE, 336–343.

Garold Stasser and William Titus. 2003. Hidden profiles: A brief history. *Psychological Inquiry* 14, 3–4 (2003), 304–313.

Martin Stettinger, Alexander Felfernig, Gerhard Leitner, and Stefan Reiterer. 2015. Counteracting anchoring effects in group decision making. In *Proceedings of the International Conference on User Modeling, Adaptation, and Personalization*. Springer, 118–130.

Cass R. Sunstein. 2006. Deliberating groups versus prediction markets (or Hayek's challenge to Habermas). *Episteme* 3, 3 (2006), 192–213.

Lawrence E. Susskind, Sarah McKearnen, and Jennifer Thomas-Lamar. 1999. *The Consensus Building Handbook: A Comprehensive Guide to Reaching Agreement*. Sage Publications.

Hamed Taheri. 2015. *Interactive Visualization to Facilitate Group Deliberations in Decision Making Processes*. Ph.D. Dissertation. University of British Columbia.

Claudia Toma and Fabrizio Butera. 2009. Hidden profiles and concealed information: Strategic information sharing and use in group decision making. *Personality and Social Psychology Bulletin* 35, 6 (2009), 793–806.

Amos Tversky and Daniel Kahneman. 1975. Judgment under uncertainty: Heuristics and biases. In *Utility, Probability, and Human Decision Making*. Springer, 141–162.

UNU. 2017. UNU.ai. (2017). Retrieved May 4, 2017 from http://www.unu.ai.

Peter C. Wason. 1960. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology* 12, 3 (1960), 129–140.

Suzanne P. Weisband. 1992. Group discussion and first advocacy effects in computer-mediated and face-to-face decision making groups. *Organizational Behavior and Human Decision Processes* 53, 3 (1992), 352–380.

Kent Wittenburg and Tommaso Turchi. 2016. Treemaps and the visual comparison of hierarchical multi-attribute data. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, 64–67.

Ludwig Wittgenstein. 2010. *Philosophical Investigations*. John Wiley & Sons.

Ji Soo Yi. 2008. *Visualized Decision Making: Development and Application of Information Visualization Techniques to Improve Decision Quality of Nursing Home Choice*. Ph.D. Dissertation. Georgia Institute of Technology.

Ji Soo Yi, Youn ah Kang, and John Stasko. 2007. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1224–1231.

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*. ACM, 686–699.