CARNEGIE MELLON UNIVERSITY

Pith: A Space for Productive Discussion

by

Sydney Zheng

advised by

Dr. Kathleen M. Carley

in the

School of Computer Science

May 2021

CARNEGIE MELLON UNIVERSITY

Abstract

advised by Dr. Kathleen M. Carley
School of Computer Science

by Sydney Zheng

In this thesis, we conceive, design, and implement the core of our experimental Pith platform for large-scale, productive discussion that encourages diverse perspectives. To encourage productivity, our system has two levels of summarization. Large-scale discussion is represented as a map of ideas, the first layer of summary. The map of ideas evolves as participants form small groups that communicate to explore the ideas. Additionally, users in a small group have a panel adjacent to their chat for them to summarize the key points of their ongoing conversation. To encourage diversity, we investigate the use of group meta-cognitive reflection mechanics on the large-scale summary and mechanics to encourage active listening in the small-scale discussion. Finally, we conduct initial testing of our system by focusing on the small discussion interface, which indicates that the interface is easy to use, that the summarization feature helps users organize their discussion and visualize its progression, and that our proposed flairs mechanic is a promising feature to encourage active listening. The second observation offers support that the large-scale summary would likewise help a large group organize and visualize their overall discussion.

Acknowledgements

I would like to first thank my advisor, Professor Kathleen Carley, director of the Center for Computational Analysis of Social and Organizational Systems (CASOS), for giving me her time and resources to guide me on this project. It was through her that I first found my love for understanding social structures.

I thank Professor Hirokazu Shirado for sharing his expertise on social science experimentation as we were creating our group studies on human subjects to test the system developed through this thesis.

I thank Professor Lynne Reder for giving us insight into the realm of hypermedia, which inspired the network organization of the Internet by which we share knowledge. I thank my systems professors, Professors Brian Railing, David Eckhardt, Seth Goldstein, Nathan Beckmann, Mahadev Satyanarayanan, Padmanabhan Pillai, and Runting Shi for giving me the knowledge to build this project.

I thank the Undergraduate Research Office for providing the Small Undergraduate Research Grant (SURG) in order for us to conduct this research.

I thank Sienna Watkins for helping setup the meetings between Professor Carley and me as well as helping us transfer the SURF funds to use on Prolific.

I thank Michael Kowalchuck for helping setup our access to the CASOS servers to store data collected from our studies.

I thank our volunteers, recruited from Prolific, for spending the time to help us conduct our experiments. Additionally, I thank our contacts on Prolific for their guidance on using the site and their help on the issues that arose during our studies.

I thank my friends for their wonderful company as I was working on this project. I would especially like to thank Maia Iyer, Oscar Dadfar, and Alex Havrilla for enlivening day to day existence during this pandemic year.

I thank my parents for their love and support, even from far away, for their patience during difficult times, and their cheer during the happy ones, for putting their all into giving me the opportunities to grow into a fulfilled being.

Last, I would like to thank Christian Broms, who is both my partner on this project and my partner in life. Inspired by our mutual fascination of experimental decentralized systems, we worked together to conceive, design, implement, and test the Pith discussion system discussed in this thesis. I concentrated on backend implementation, the test setup, and the experiments, while he concentrated on frontend implementation and infrastructure.

Contents

Chapter 1

Introduction

Conversation is the means by which we understand concepts, experiences, and even our own social identity [1]. Our understanding of the world is often formed by what we can have a conversation with. Through others near us, we come to understand local experiences. Through books, we come to understand in depth experiences of the past or experiences within a particular domain. Through television, we come to understand the experiences of our mainstream culture. Through the internet, we have the potential to access experiences beyond our own locality, beyond any established topic, beyond the norm.

The internet has endowed us with the potential of creating a place for us to understand the perspectives of those far away and those very different from us. We argue that we must realize this potential in order to solve our greatest challenges in the decades to come. As the world has come closer together throughout the past centuries, we have arrived at a point where the lives of "people here" have become tightly integrated with the lives of "people there". The decisions we make regarding climate change, transportation, commerce, human rights, and so on often not only affect ourselves but people far from where we are. The decisions they make not only affect themselves but also us. Thus, in order to address these challenges, we need the diverse perspectives of people who are affected and people who may have potential solutions. We need a system that allows for productive collaboration across the globe.

We currently rely on government, sometimes businesses, other organizations, or academia, to come up with the solutions that address our needs. Therefore, discussion on these issues are often siloed into their respective sectors. Shum et al. imagine a "Global Participatory Platform" [2], an ecosystem of discussion that opens up boundaries not only between these sectors but also between these sectors and the public, between those who are affected and those who can plan and act.

In line with this vision, this thesis describes the creation of a discussion system, Pith, that aims to be a first step in designing an experimental space to allow many different people to come together, share experiences, and produce solutions. It aims to do so in a way that considers the diverse perspectives of participants, with the intent of overcoming polarization and echo

chamber phenomena that have dominated traditional social spaces, most notably social media.

In the Background chapter, we describe existing, similar systems that have allowed us to formulate a set of desired attributes for the ideal discussion platform. In the Design chapter, we highlight the key goals for our own system and how we address each goal. In the System chapter, we describe the implementation of Pith. In the Experiments chapter, we describe what our initial set of tests reveal about the potential of the system. Finally, in the Conclusion chapter, we outline the current limitations and the future work. Even after this thesis, we hope to further Pith in becoming a working productive, large-scale discussion platform that encourages diverse perspectives.

Chapter 2

Background

We describe the different classes of systems that yielded insights about the form of a large-scale discussion system. We then highlight the attributes we believe desirable for such an ideal discussion system and summarize how each existing system incorporates these attributes through a table of advantages and disadvantages.

2.1 Classes of Related Systems

2.1.1 Open Platform

Open platforms are designed to allow a mass of people to contribute their ideas or thoughts to some issue. We further divide this category into e-democracies, idea spaces, and debate spaces.

2.1.1.1 E-Democracy

The goal of an e-democracy space is usually to allow citizens to work together to achieve goals that benefit their community.

DemocracyOS is a canonical example of how an online space intersects with the decision-making process in real life [3]. It is a digital system used by the Net Party of Argentina to communicate with their constituents. Policymakers present proposals to citizens, who can then provide feedback on them. In addition, votes can be held on various issues. However, DemocracyOS requires the citizen-user to have high technical literacy in order to use it.

Australia and the United Kingdom have created a space for citizens to provide government feedback about their proposals [4, 5]. In the United States, former president Barack Obama created a space to allow citizens to share stories with the government about people in their community doing great work that the government could reach out to.

Another example is E-Democracy.org, which is a forum space that allows citizens in the same geographic community to communicate and potentially organize local efforts, such as electronics

recycling [6]. Unlike the previous examples, it is not limited to one country, allowing for a general framework for citizen collaboration.

Each of these systems offer great opportunities for individuals to get involved with their community or country's needs. However, in DemocracyOS's case, decisions are arrived at mostly through a voting mechanic rather than discussion. The country-based spaces are mostly oriented on citizen feedback on existing proposals. While this method provides a good foundation for a plan of action, it diminishes the opportunity for citizens from a wide spectrum of backgrounds to talk to each other and collectively imagine their own solutions. Though E-Democracy.org proposes such a place, its forum-based interface is arguably difficult for people to organize a coherent plan of action.

2.1.1.2 Idea Space

Somewhat similar to the e-democracy systems are idea spaces. An idea space crowd sources ideas from a mass of users on some topic. Because there may be many ideas, there is usually some mechanism to make the "best" most visible.

Salganik and Levy's All Our Ideas platform is one such example of an idea space [7, 8]. Anyone can create a question with an initial set of answers. This question is presented to the masses, who can choose which of the given answers they agree with or create a new answer if they desire. Ideas with more votes have higher visibility. The advantage of this system, as Salganik and Levy describe in their 2015 paper, is that experts can bring in their own insight when creating the questions by forming the initial set of ideas. Thus, users seeing the question for the first time have a framework of ideas to work off of. However, the many users can bring their creativity or diverse experiences to come up with additional, potential better ideas.

While All Our Ideas asynchronously draws idea on a topic, Synthetron allows a mass of participants to work together synchronously to brainstorm ideas, usually for an hour [9]. An organization may intend to brainstorm with all its members, a company may want to get feedback from its customers, or a policymaker may intend to hear the ideas of their constituents. A moderator presents a set of questions, and for each question, the participants can anonymously contribute ideas into a chat-like system. Once existing ideas are presented, participants can augment them by providing additional commentary. Then for each idea that has been presented, participants are expected to weigh how much they agree or disagree with each idea. After this process, the average score is presented for each idea, indicating which how accepted they are in comparison to each other. The advantage of the system is that it can theoretically let thousands of people participate at once, and the group as a whole can see which ideas are more desirable. In addition, there is a clear product from this process, which is a set of ideas that the organization can use to address each question. Because the process is only an hour, it is clear when the product will be available.

However, both systems are more intended for brainstorming than for a discussion. Thus, the intention is usually to derive insight from the "wisdom of the crowds" and their general views.

Less so is the intention to inspire participants to question each other, share experiences, or to architect a solution or response more complicated than a set of ideas.

2.1.1.3 Debate Space

Debate spaces are similar to idea spaces in that they can use the wisdom of the crowds to draw in a vast array of ideas. However, they are different in that their focus is usually to have participants consider if they support or oppose some view through the delivery of arguments on both sides.

Kialo allows a mass of participants to asynchronously participate in a debate [10]. People argue for or against some initial statement, for example, "All humans should be vegan." Each argument, either on the agreement or opposition side, is a sub-statement that can be further deliberated, for or against. Thus, Kialo organizes each topic as an argument tree. People can vote for which arguments they like the most on the agreement or opposition side. The top arguments appear closer to top, where users are more likely to see them. Debatepedia and DebateGraph are other examples with a similar argument tree structure [11, 12].

Debates can help people quickly understand the common views on both sides of an issue. To some extent, they encourage critical thinking in that a person on one side can examine the arguments of the other side. However, they do this at the cost of polarizing the issue. Because arguments are supposed to be on one side or the other, there is less nuance for viewpoints between.

Debate.org is perhaps the most well-known example of such a space [13]. Its main feature is different from the others in that it focuses on a debate between two people. The debate has several rounds, where in each round, each of the two participants makes an argument or response. Other participants can watch the progression of the debate, comment, and at the end, vote who should win. Durmus and Cardie determined that factors beyond the arguments themselves predicted which of the participants will win the vote [14]. Namely, debaters who used the first person pronoun, i.e. "I", were more likely to succeed. They hypothesized this was due to the debater integrating personal anecdotes into their argument, suggesting the power of more subjective content in convincing others.

2.1.2 Social Media

Here, "social media" is used to mean platforms with the primary purpose of allowing a mass of people to communicate in a more casual setting. They often act as a common discussion space, allowing almost anyone to express their view.

A common concern of social media giants, such as Twitter, is their tendency to polarize issues and create echo chambers. Users are more likely to see content that they agree with. In addition, the most seen content is typically the most popular content, which has been upvoted or liked the most. This greatly decreases the relevance of many other perspectives to the common

consciousness, despite their potential contributions. However, these platforms have a plethora of features that are worth considering and a wealth of social behaviors to derive insights from about how people communicate online.

Twitter's main purpose is to allow individuals to broadcast information to their followers and to receive information from those they follow [15]. Tweets are presented as short snippets of information, which can often convey news, some interesting link, or an opinion. In contrast to idea spaces, content is primarily organized in relation to each person rather than to some set issue. However, Twitter does have a way to organize information in topic-communities. Posters can assign hashtags to their tweets, such that the tweet is searchable under each hashtag. A hashtag can reveal bodies of ideas people have been or are currently interested in. As such, it could be argued that a hashtag is a point around which discussion collects, such as the #metoo hashtag, which represents the Me Too movement for addressing sexual abuse, or the #blm hashtag, for the Black Lives Matter movement for addressing the rights of Black people. An advantage of Twitter is how an idea can spread, from community from community, instead of being siloed to some particular topic or section. People in one community can share the tweet with people from another community, thereby allowing a great confluence of different ideas.

Reddit is another noteworthy example of social media [16]. Reddit provides a medium between having clearly divisible topics, like in the open platforms, and having a cloud of ideas, like in Twitter. It does this through the use of subreddits, which are communities with some focus. Some of these are topics, such as r/datascience for data science, or r/nutrition for nutrition. However, some are more open-ended, such as r/askmen, which is a place people can discuss a wide variety of topics, such as health, relationships, sex, career, entertainment, life centered on the male experience. Each subreddit is a feed of posts, where each post focuses on whatever the author desires. A post can be a question, a piece of news or media, a writing prompt, and so on. Under each post, discussion is organized in a comment tree, where the most upvoted comments appear closer to the top. Each community has its own culture, outlined by a series of guidelines. In addition, each subreddit has moderators to ensure that users are respecting the rules and are nonabusive to each other. One of the features we were particularly interested in were flairs. Users could flair some trait of themselves, such as individuals flairing themselves as "Male" or "Female" in the r/askmen community. Posts could be flaired as "NSFW" (not safe for work) for example. We use the ideas of flairs in a somewhat different way in our system.

Similar to Twitter is Mastodon, created by Eugen Rochko, which was designed to address several concerns of social media as online spaces [17]. Typically, social media is heavily monetized through ads and is owned by a corporation. Twitter has standard moderation rules that may not be desirable to all users. Mastodon allows the formation of communities that do not have to be hosted by a corporation's central server but through decentralized servers, spun up by anyone. Each group can formulate their own culture by which to moderate themselves by.

Mattermore is a more recent example of social media, useful among researchers [18]. It allows diverse experts to form a community around some salient topic, such as climate change or COVID, and work together to contribute understanding. Because of this, it serves as a modern, working example of a discussion platform, even it is targeted to the academic community.

A more niche example of a social media-like site is Special Fish [19], which presents a colorful disorganized board with little "fish", where every fish is a profile of some individual. Whereas people typically navigate around sites like Twitter and Reddit through search, Special Fish illustrates an alternative to finding new content, in a more exploratory way.

2.1.3 Group Communication

Whereas the former categories of systems have content meant to be shared by the general population, the following systems highlight the main ways a closed group, such as an organization or close set of friends, have discussions online.

2.1.3.1 Thread-Based

Thread-based systems allow users to communicate through text, typically in a stream-like a chat.

Facebook's Messenger is one of the most commonly used methods of chatting with others [20]. A group of people have a chat space, by which anyone in the group can make a post that everyone else can read. The conversations on Messenger are typically more casual in nature, and can inspire people to share their everyday experiences. Messenger has some additional features worth noting. It has a basic search functionality to search for different posts by keywords. There is the ability to reply-to specific posts just as in a real, group conversation, one person can turn to someone who made a particular comment and address that comment. It also allows for a mechanism of non-verbal communication beyond the text through emojis, which convey salient emotions or other reactions.

Slack [21] and Discord [22] are also messaging services, where the former is more often used for work and the latter more often used for gaming or casual hangouts. The two have the essential features of Messenger along with an additional main feature, called channels. Instead of a group of people using only one space to have a conversation, they can have different spaces for different conversations. For example, having a "fun" channel for fun posts, a "customer" channel to talk to clients, and a "meeting" channel to arrange meetings. In addition, the two allow pinning, such that chosen, essential messages are easily accessible on a sideboard.

Another intriguing form of chat is Quill, which further extends the concept of channels in order to organize conversation [23]. Conversations can be organized in a folder-like structure to further increase the topicality of a thread. Even the course of a focused conversation, the topic may change. Users can take the posts that indicate the change in topic and put it in a new thread. In addition, priority messages can be used to distinguish certain messages from the others.

While these different messaging systems all allow for natural conversation, they lack a means by which people can organize their ideas. Even Quill, which attempts to preserve the focus of a conversation would typically have a high noise-to-signal ratio such that most posts are general conversation and only a few posts express some salient idea. Further organization or

preservation of ideas is desirable in the case where people are conversing to develop a framework of ideas, such as a solution to some problem.

In addition, conversation among a larger number of people becomes increasingly difficult, especially if they are in the same thread. On average, each person's posts only appear sporadically within the overall stream of posts.

Facebook's Workplace somewhat addresses both concerns through the use of communities [24]. Within an organization, such as a business, there may be various formal sub-organizations or even other groups, such as those based on interest that cross over the natural tree-like structure of most organizations. Each group has their own community page by which a feed of posts is continuously generated. People can report news or other ideas of interest in these community groups. Just like the above systems, it allows for one-on-one or group chats with more casual conversation separate from the community feeds. However, the posts in each community feed are typically more salient than the posts in the chats. Thus, the feed generally represents the space of ideas that are more noteworthy to a group. However, because the structure of a community's ideas is a feed, it is still difficult to build more coherent solutions on this platform alone.

2.1.3.2 Meeting-Based

Many of the aforementioned systems do have a video component but emphasize text. However, the following systems emphasize video. Video offers greater synchrony than text-messaging, so conversation often feels more engaging.

The canonical video-based meeting system is Zoom, where a group of participants can see each other and talk together [25]. A challenge of video-based systems is turn-taking. Zoom allows participants to "raise their hands", such that if the meeting was being moderated, the moderator can call out who can talk next. It also allows for a simple set of emoticons, like clapping hands or thumbs-up. Another feature of special interest are break-out rooms, which allow a large group to be broken into smaller groups and later, to return back to the large group. Smaller groups provide individuals with more opportunity to speak, even if it is to a smaller audience. However, one can imagine that the ability to temporarily divide a large group into smaller ones allows for efficient work distribution of certain problems. For example, to plan for a party, a large group can break into small groups where one small group in a breakout room can select the food, another small group can choose decorations, and a third small group can decide on invitees.

Airmeet [26], Gather [27], and MakeSpace [28] further the concept of creating small groups. Airmeet allows users to seat themselves at virtual tables, where each table holds up to eight people or so, and have a small conversation through video-chat. Gather lets people direct avators through an RPG map that simulates a physical space, such that there may be many people on the map but each person can only video-chat those who appear close to them. MakeSpace allows people to divide themselves into rooms of their own making.

Meeting spaces sometimes have a means of preserving ideas. Zoom offers a small chat space people typically use to drop media or links or to ask questions. Some video-based system have an in-built whiteboard feature to allow users to further organize their ideas, such as in a visual map. Alternatively, video-callers can use external systems like Excalidraw [29] or Miro [30] to structure their ideas. The former is a simple system to create text in shapes with arrows, and the latter offers more sophisticated templates for organizing ideas, such as a mind-map and Agile board.

2.1.4 Knowledge Storage and Information

The previous category of systems typically lack the functionality to store knowledge that arises from conversation in a coherent way. The following systems provide different methods of organizing information, typically at the cost of a group's ability to have a natural conversation.

A common tension arises in how one organizes information. Humans naturally organize knowledge in two ways. The first is categorically or hierarchically–dogs are a kind of mammal which are a kind of animal. This mental model is typically represented by a tree structure, such as the taxonomy of life, which organizes life according to a tree of domain, kingdom, phylum, class, order, family, genus, and species. However, just as there are similarities in two fairly unrelated animals in this tree of life, such as that bats and birds can both fly, humans connect ideas that are first may seem different. Thus describes our second mode of thinking, associative thinking, which taken to its extreme, has ideas organized in a highly-connected network. The balance between a hierarchical organization versus a more flexible network organization one often has to be considered in developing a system to organize knowledge.

2.1.4.1 Hypermedia

Hypermedia describes the ways of organizing digital information in a nonlinear fashion. Indeed, it was through hypermedia that the notion of web pages and links first came about to describe the architecture of the World Wide Web itself.

The Hypertext Abstraction Machine (HAM) presented in 1987, ZOG developed here at Carnegie Mellon in 1984, its successor, KMS in 1988, and Notecards in 1988 all generally propose a network model of organizing units of information, such that these units have links to other units [31–34]. In the proposal for KMS, Akscyn, McCracken and Yoder make a wonderful metaphor: that the described database of ideas is like universe of connected spaces, where a pilot can zoom from one place to another. However, Halasz in Notecards makes a crucial observation about a purely network structure: the lack of some kind of hierarchy is off-putting to users, who often have difficulty in finding information in a more systematic way. They cannot "zoom in and out" of the information–they cannot see a high-level description of it and then delve into the details they desire in a way they can better do with a well-organized folder system. Additionally, Halasz notes the need for search beyond the traversal of links, a point which is more readily addressed in modern technology; search engines allow us to query web pages through keywords, search

indices allow us to query posts in a thread. Halasz also anticipates the need for collaborative work, that the creation and organization of information should be performed by more than an individual in many cases.

2.1.4.2 Argument Map

The argument map is similar to the argument trees that were used for mass debate in that they often have a tree structure at their core and people generally try to argue for or against positions. However, they typically have more structure in order to decide upon potential solutions, which makes them both more flexible and often harder to use.

Klein proposed the creation of the large-scale argument mapping system, Collaboratorium, in 2007 [35]. His motivation was to design a system for a large number of people to address complex issues, in contrast to the systems typically built for small-scale groups. He argued that crowd-sourcing systems like idea spaces often resulted in the duplication of ideas and the general lack of structure surrounding ideas. He thus suggested a way to map out ideas through a conversation with three main entities: *issues* to raise questions, *options* to propose solutions, and *arguments* to deliberate each option. However, while natural conversation as enabled by the group communication systems of the previous category allow for information to be organized in time, he considered that ideas should be more organized more by topic or their function in relation to other ideas.

In 2009, he along with Iandoli and Zollo proposed a new system called the Deliberatorium that formalized their designs on argument maps [36, 37]. Their main question was whether the system could scale to a large number of people. They emphasized that a discussion platform was more suitable than a wiki or a forum for large-scale discussion because it could better allow individuals to come to an understanding and thereby minimize the effects of polarization. Alongside the argument map, they included auxiliary tools for chatting and a forum space. In addition, they included search to prevent people from authoring duplicate content, and a moderation system that oversaw the organization of the argument map. Only moderators had certain privileges, such as editing posts. They began initial tests on their system, which indicated that indeed the argument map system allowed a group of people to cover knowledge on some topic. However, there was more content generation, creation of new posts, than content reflection, commenting on existing posts.

In 2010, the three authors along with Gürkan performed more empirical anlysis on the Deliberatorium system, thereby orchestrating what was likely the largest argument map at the time with 5,000 posts [38]. Their main conclusion supported their results from 2009, in that the argument map allowed a group to quickly cover the knowledge on some topic. However, they noted the extensive need for moderators, estimating that about 5-10% of those that generated content needed to ensure the organization of the content within the argument map. Additionally, they discovered that a lot of communication occurred within chat rooms or forums, indicating that more natural conversation of some kind was necessary to architect the structure. As before, they noted that people would rather create content than evaluate content, and so further considered

systems such as rating posts or individuals. Indeed, such a problem could indicate individuals are more interested in "speaking" rather than "active listening".

In 2012, Klein explored means to further prevent idea duplication and help people identify spaces of the argument map they could be interested and sufficiently able in contributing to [39]. Especially as the argument map grew larger, tools needed to be made to help draw people's attention to the most productive place. He proposed four metrics to judge different "exceptions", or mechanics that prevented the large group from converging onto a solution and that were places people could best contribute to. The first exception was an immature issue, which often does not have enough exploration of the solution space to reach convergence. The second was groupthink, whereby people focus mostly on just a few popular solutions instead of more carefully considering the alternatives. The third was balkanization, by which a community is segmented and forms different camps supporting different solutions, thus preventing convergence. The fourth was non-independent ratings, generally the phenomena that people tend to like ideas that are more popular.

Collaboratorium and Deliberatorium serve as outstanding predecessors of a large-scale discussion system. However, they emphasize objective viewpoints and have a rather rigid structure, which is generally a tree. However subjective viewpoints can promote a sense of stakes or empathy about the issue. The latter suggests that ideas that cut across topics are not so common. However such "bridging" ideas may be highly useful in drawing knowledge from one domain to another.

Other argument-mapping system have been proposed, such as Compendium, which can help a group map a thought-process as a dialog of why certain decisions are useful are not [40]. Concepts are connected with propositions or verbs. While this approach clarifies the structure of a though-process, it does not appear to have the "zoom" property of Collaboratorium or Deliberation, where ideas can be collapsed into some topic or otherwise summarized to help people navigate through them.

Krauthoff et al. build on the work of the canonical argument map with a topic, position, argument structure, but help users more easily follow the structure [41]. They proposed a system called D-BAS which guides users through an initial multiple choice asking how they wish to respond to a post: whether they agree or disagree with its position, and whether they agree with its supporting argumentation. Users can then make their text-based post with the proper connection to the former post.

The general structure of an argument map has also been designed for use by corporations with the Structured Evidential Argumentation System [42], to help them record their knowledge from the past and present to apply to the future, essentially serving as a database of ideas and arguments for different issues. Interestingly, it also addresses issues of access rights in the database.

An argument map structure was also one of the key components of the Climate CoLab, introduced by Introne et al. in 2011 to serve as a global online forum for netizens to discuss climate change [43].

The different applications of argument maps in these contexts suggest their general structure is especially useful to organized, large groups.

2.1.4.3 Idea Map

Akin to the argument map is the idea map, also known as the mind map. Compared to the argument map, the idea map has less intrinsic structure, forgoing the need to classify posts as position, topic, argument or other similar schemas. In addition, idea maps usually follow a more network structure rather than tree structure. A potential advantage of the looser structure of the network, as discussed before with Special Fish, is that its more cluttered appearance may encourage exploratory searching, also suggested by Janiszewski in 1998 [44].

A simple, niche example is Kinopio, which allows a group of people to create a mind map [45]. Another example is Knowledge Forum, which is a tool used by groups to present ideas in a scaffold in such a way that others can build on it [46]. For example, students can propose the structure of a group essay that takes from different sources and get feedback on their ideas from each other and their teacher. It also has an intriguing feature called "views", which allows the same idea to be seen in various visual contexts, exposing the different network relations it is a part of.

2.1.4.4 Recursive List

Whereas idea maps can be considered the extreme of network organization, recursive lists are the extreme of a hierarchical organization.

Workflowy and Dynalist offer a recursive list product, where a user can create an infinitely long list, such that each bullet point can have infinitely many depths of descendants [47, 48]. Naturally, this structure can be used to organize sections and sub-sections, such as for a document. However, this structure can also be used to create arbitrary lists, where each item has more detail. For example, a to-do list, where under each to-do bullet is a description of the details of the task or the eventual execution of the task.

In some senses, even a reply-to thread, like a series of comments, is a recursive list, yet the function of the aforementioned products is to preserve salient ideas rather than posts in a time-based conversation.

2.1.4.5 Document or Wiki

As Klein, Iandoli and Zollo refer to in their 2009 paper, wikis can be considered another form of large-scale discussion, in that they preserve the knowledge of some subject [36]. The wiki represents the state of a community's knowledge on some topic, and changes to reflect a community's progressing understanding of the topic. Wikis are typically divided into pages, where

each page is of some topic. The page is organized like a document with sections and subsections and may include internal links to other pages in the wiki or external links. Typically a few authors of the community perform most of the writing on the topic, though others can theoretically edit it. However, a common disadvantage is the difficulty in expressing complex topics, especially contentious topics in how to portray the different viewpoints. This is related to another disadvantage of wikis in that they typically suffer from an ability to express more personal experiences, emotions, and anecdotes which can communicate the weight of an issue.

Wikipedia is perhaps the best known wiki, serving as the canonical knowledge base of the internet [49]. Pages are linked together in a network of information, though on each page, information is presented in topical sections. Theoretically anyone can edit any page. Wikipedia has a set of guidelines to shape the culture of the community that uses it: that its purpose is only to be a source of information, that the writing should be neutral, that it should be free, that editors should respect each other, and that otherwise there are no firm rules. The authors are for the most part equal, but those with good standing in the community can become administrators. Thus, it has a somewhat meritocratic social system of assigning who has more authority. Though it allows some discussion, it is typically minimal, focused on the issues of a wiki-page and how to address them.

There are different means by which communities can create their own wiki or means of outlining ideas on some topic. The simplest method is to use a shared document, such as GoogleDocs [50]. A more sophisticated option would be to use Roam Research or Notion [51, 52]. The former was designed to help facilitate associative thinking, to help researchers draw resources from different sources and to form connections. Their +Roam browser extension tool allows users to simply copy text from any webpage and easily copy and cite the text in their Roam Research database. Both Roam Research and Notion allow users to organize their information in pages, which can be composed of raw text, headers, and lists. However, both allow information to be organized in a network through the use of backlinks. A user can reference another page in the system, and the other page will display that it has been referenced by the first page, thereby creating a bidirectional link. Roam Research also allows users to view this network explicitly in a node-link visualization–however, as Halasz remarks in NoteCards, a network visualization is difficult to navigate, and indeed this is the case with the network visualization.

2.1.5 Post-Analysis

Also of interest are systems that allow post-analysis of discussions or information, such as to reveal insights about its structure.

In particular, several efforts have been made to summarize information on the fly. Zhang et al. propose the Pegasus system, which uses machine learning to perform summarization on some text [53]. A more concerted effort to summarize conversations is Wikum [54, 55]. In their 2017 paper, Zhang, Verou and Karger note the difficulty of "digesting" large discussions because of their lack of summarization. They use tags and automatic clustering to suggest which group of posts, comments, or threads can be summarized together.

Another note-worthy form of analysis is to capture the space of the discussion in a more condensed form. VOSViewer is a tool to allow one to study bibliometric networks, its authors, topics, publishers, and institutions, who cites who, who publishes who. These networks act as a kind of summary of the research space [56]. Horn conducted an analysis of discussion in history up to current times, as represented by how documents cited each other [57]. To present a high-level representation of the discussion, he clumped together similar arguments. His goal in creating this discussion network was to show how humanity has worked together on major questions over time, cutting across disparate fields of inquiry.

ORA is a research tool developed here at Carnegie Mellon by the Center for Computational Analysis of Social and Organizational Systems (CASOS) to study networks not only in the research sphere but in the everyday sphere, through analyzing social media [58, 59]. For example, a researcher can study how people may generally think about the COVID pandemic by analyzing tweet conversations. One can identify communities based upon who follows who and what the hashtag or topic of their tweets tend to be. ORA allows users to see the network in 2-dimensional space, and has a hyperbolic view that allows one to "zoom in" on a section of the network to study in greater detail while still seeing the rest of the network on the periphery.

2.2 Desired Attributes

After examining the aforementioned systems, we considered what form a large-scale discussion platform should take.

Town and Herbsleb suggested that a large-scale discussion platform should have the following characteristics: it should attract contributions, the deliberation content should be navigable, the output should be reasonably useful, it should encourage quality in content, and it should promote wide-scale adoption [60]. In addition, they suggested a number of other attributes. For organization, they suggested that it should minimize duplication, that it should emphasize topic or logic over a chronological organization, and that its structure should be more organic than strict. They encourage the ability to reference, such as through linking to external resources and linking to internal content through hyperlink. They recommend mechanisms to allow a divide and conquer of work, such as splitting a topic into subtopics. For quality control, they suggested a rating or reputation system for posts or authors. To encourage adoption, they suggest that it should be interoperable with other systems and supplement rather than usurp other systems.

In addition, Macnamara argued that an interface must be designed to encourage active listening, which is necessary among the constituents of a group to prevent fragmentation of knowledge and polarization [61]. As demonstrated in the experiments with Deliberatorium [38], people have a tendency to generate content, to "speak", rather to reflect on other's content, or to "listen". Thus he proposed the following means: to have some background material on a topic so that participants have familiarity with the essentials, for there to be moderation to prevent undesired social behaviors such as the use of ad hominem attacks, to have an acknowledgment function of content, to make content navigable such as through categorization with headers and topics and summaries. Some of these suggestions are implemented by existing systems. For example,

Twitter [15] and Reddit [16] have extensive moderation systems. Many platforms, social media and otherwise, have some form of acknowledgment that often centers around likes or upvotes, subscription or following, favoriting or sharing. Content is somewhat navigable through search, and for argument maps, through a tree structure. However, his suggestions with summarization in particular are not common outside of systems like argument maps.

We thus formulated the following set of desired attributes for an ideal large-scale discussion platform based on the aforementioned recommendations and our own observations of the various systems.

2.2.1 Natural

The first quality is that participants should be able to communicate in a natural way, in a way that feels like conversation. Conversation allows individuals opportunities to ask questions, share their experiences, and convey knowledge. In addition, natural conversation can humanize participants to each other, imparting a greater weight to what is being said.

2.2.2 Objective and Subjective

As mentioned several times in relation to systems like debate spaces and argument maps, subjective content is just as critical to discussion as objective content. Objective knowledge is useful for people to understand empirical reality and how to achieve practical outcomes on the world. However, subjective knowledge, the stuff of anecdotes, beliefs, and values, allows individuals to feel the weight of each other's conviction, of each other's experience, and thus of each other's voice.

2.2.3 Minimizing Polarization

The discussion should not promote people to take sides and robbing them of the opportunity to consider the nuance of human experience and potential solutions that exist beyond the side-based framework. The system should avoid fracturing the community as a whole, such as by limiting what people are exposed to, even for the sake of their perceived interest.

2.2.4 Common Understanding

Participants in a discussion should be aware of and understand what has been brought up in a discussion as they continue to progress. This prevents duplication of content and allows them to focus on what hasn't been addressed.

A common understanding can be created through some summary of the shared structure. A summary may appear in the form of a condensed piece of text or a condensed document, or even in a descriptive network of ideas that outline the scope of the discussion.

A summary also addresses the issue for creating a practical outcome of the discussion. The summary can be a working product, that at the end, offers the exploration of an issue or outlines a solution to a problem.

2.2.5 Storage, Navigable, Queryable

Following the former point, the discussion should output knowledge or ideas that can be stored and later retrieved as participants desire to explore or search across it. The organization of ideas should encourage exploration as well as helping people find exactly what they are looking for.

2.2.6 Sense of Scale

Should the visited idea space grow, there should be some way to zoom out and see ideas at a high-level and zoom in to some particular detail. A common way of organizing information to offer a sense of scale is through a hierarchy, where nodes closer to the root are more "zoomed-out" and nodes closer to the leaves are more "zoomed-in". Even a network structure can be contracted–nodes can be grouped into clusters, such that at lower resolutions, the cluster appears to be a node, and at higher resolutions, a node appears to be a cluster of nodes. However, there could be other ways to encourage a sense of scale. As mentioned before, ORA uses a hyperbolic view to allow a user analyzing a network keep in sight the whole network and also focus on a particular detail at the same time [59].

2.2.7 Bridgable

The system should allow ideas to be easily referenced, should allow communities to be bridged. This is more typical of a network structure of information as compared to a more rigid tree structure.

2.2.8 Encouraging Diversity of Ideas

A greater diversity of ideas leads to a more thorough exploration of some topic. Participants should be encouraged to share their perspective feel like an active, considered part of the community. Mechanisms against this are popularity metrics, and to some extent, systems like rating and reputation.

2.2.9 Minimizing Social Hierarchy

The system should encourage self-moderation rather than external moderation wherever possible. This is generally to encourage a more democratic mindset, somewhat in line with the former attribute.

Formal moderators often have the following goals: to organize content that users generate, to influence the high-level course of the discussion or content generation, and to regulate user behavior. We are especially interested in how users can address the first and second goal themselves through group meta-cognition, which will be described in the next chapter. The last goal, we defer for greater consideration later.

2.2.10 Low Barrier of Entry

As mentioned by Town and Herbsleb [60], a low barrier of entry attracts contributions, especially a diverse range of contributions. Ideally, a system should be easy to use, reasonably enjoyable, and integratable with the rest of an individual's activities. The ideal platform is not isolated but coexists with the other efforts of human inquiry and progress.

2.3 Comparison

In the following table, we summarize the advantages and disadvantages of key products or classes of product in consideration of the desired attributes of a discussion platform. Note that some listed disadvantages may be considered advantages in other contexts.

Category	**System**	**Advantages**	**Disadvantages**
Open Platform	E-Democracy forum [6]	Encourages citizen action or productivity. Open community of ideas. Forum allows semblance of natural conversation, sharing of objective or subjective viewpoints. Some sense of scale with information organized by geography.	Generally speaking, lacks sense of scale and information structure.
Open Platform	DemocracyOS [3]	Encourages citizen action or productivity. Citizens have some sat in modifying a common document, the policy proposal.	Primarily vote-based rather than discussion-based. High barrier of entry, requires technical proficiency. Policy-maker has final say.
Open Platform	All Our Ideas [7]	Integrates expert knowledge, crowdsources a diverse array of ideas. Easy to use, allows some exploration of different viewpoints.	Ideas are one-off rather than part of a working solution. Little communication to question other points or gain additional context.

Open Platform	Synthetron [9]	Allows synchronous, large-scale communication and efficient brainstorming.	Requires users to rate proposed ideas, even if they have little context. Limited mechanisms for back-and-forth discussion and means of architecting complex solutions.
Open Platform	Debate [10–13],	Open community of ideas. Encourages objective thinking. Can be like natural conversation depending on format.	Polarizes topic. Ideas often evaluated through votes or weighting rather than further discussion.
Social Media	Twitter [15]	Open community of ideas. Spread of ideas across communities through sharing. Exploration of ideas through search and hashtags.	Doesn't encourage depth of ideas through character limit. Has been known to create echo chambers, encouraging polarization. Uses popularity metrics to prioritize certain authors or viewpoints. Requires moderators. Unlike natural conversation.
Social Media	Reddit [16]	Open community of ideas. Comment thread allows for some semblance of natural conversation, sharing of objective or subjective viewpoints. A sub-reddit acts as a knowledge base for ideas.	Within sub-reddit community, often a lack of some sense of scale. Requires moderators. Uses popularity metrics to prioritize certain ideas.
Group Communication	Text-Messaging or Chat [20]	Natural communication. Encourages sharing subjective, personal experience.	Lack of concise, easily navigable or searchable storage structure of ideas.
Group Communication	Slack [21] and Discord [22]	Natural communication. Based on context, encourages objective and subjective viewpoints. Pinning for idea storage. Reply-to for cross-reference. Some structure with channels.	Pinning is a weak storage method. Often difficult to get gist of discussion.

Group Communication	Quill [23]	Natural communication. Objective and subjective viewpoints. Has some sense of scale with tree-like channel system. Has pinning, reply-to mechanisms.	As a chat system, still has high noise-to-signal ratio. No concise place to store and retrieve information.
Group Communication	Workplace [24]	Has communities with feed which each have a somewhat concise space of knowledge. Has chat for natural conversation.	Somewhat difficult to find information in community feed. Lacks sense of scale of information.
Knowledge Storage and Organization	Hypermedia or Personal Web-Pages [31–34, 62]	Network storage. Can have sense of scale based on organization of information.	Difficult to communicate with others.
Knowledge Storage and Organization	Argument Map [35–37, 40, 41]	Well-organized exploration of a topic by a group. Primarily tree structures are used. Encourages objective thinking.	Structure requires high learning curve. Does not encourage sharing subjective, personal viewpoints. May encourage some polarized thinking, by having arguments for and against positions. Often requires a high content moderator to user ratio. Difficult to communicate in a natural way.
Knowledge Storage and Organization	Idea Maps [45, 46]	Well-organized exploration of a topic by a group. Encourages interconnected thinking.	Difficult to communicate in a natural way. Depending on structure, may lack sense of scale, limiting complexity of idea-space.
Knowledge Storage and Organization	Recursive Lists [47, 48]	Structured organization of ideas by a group. Clear sense of scale through tree structure.	More difficult to form links. Difficult to communicate in a natural way.

Knowledge Storage and Organization	Wikipedia [49]	Common knowledge for people worldwide. Encourages objective thinking. High interconnectivity of topics via links.	Fairly high barrier of entry. Some social hierarchy based on meritocracy. Does not encourage subjective thinking, personal experience, or natural conversation. Does not always encapsulate disparate viewpoints.
Knowledge Storage and Organization	Roam Research [51] and Notion [52]	Allows group to form shared space of information with tree-like and graph-like properties.	Does not allow for natural conversation.

Chapter 3

Design

3.1 High-Level Design Goals

Our three main goals in designing Pith were:

1. **Scale**: How can we allow a large group organize to communicate among its members naturally?
2. **Productivity**: How can we help the group reach a meaningful product (i.e. summary) at the end of their discussion?
3. **Diversity**: How can we help elevate the voice of each individual, such that each has opportunities to influence the course of discussion? How can ideas (i.e. minority viewpoints) get the attention they deserve?

3.1.1 Scale

As discussed before, we came to the conclusion that some form of direct communication was necessary. However, it is difficult for a large group to communicate among itself. The naive method would be for all of them to discuss in a large group at once, such as through a mass-text. However, this solution is ineffective in maintaining a thread of discussion, and grants each individual on average little opportunity to share their view. This naive solution is therefore not scalable.

Each individual instead needs ample opportunity to present their view in the discussion. This is far more natural in a small discussion. The thread of the small discussion is easier to follow with fewer contributors. In our platform, the small discussion is carried out through a simple chat.

Thus, our first idea was to progress the overall discussion through many small discussions. With this, two questions arose:

1. How do the small discussions arise?

2. How do the small discussions progress the overall, large discussion?

3.1.2 Productivity

Our second main idea, which addresses productivity as well as the above questions, was to use a "shared space" that everyone can see and modify within the small groups. It can be thought of as a document, similar to the wiki or an idea map, summarizing the ideas shared across the many small discussions. Participants organize themselves into small discussions based on points in the document, and can modify the document based on the output of their small discussion. The state of the document therefore represents the state of the overall discussion. In addition, the state of the document at the end serves as the final "product" of the discussion.

3.1.3 Diversity

Our third main idea was to use group meta-cognition. Essentially, there should be mechanisms by which people can reflect on the overall state of the discussion and influence its progress. Generally, how can the individual or a small group influence what the overall group should discuss and when? This might be especially important when an individual or a group of people believe an idea has not been addressed as much as it should have been. Group meta-cognitive features are intended to help participants partake in further reflection of what has been said and to identify what needs to be discussed further.

We had two main phases of design: the first was in the fall semester, and the second was in the spring semester. The goal of the first phase was to design an "alpha" prototype of the system that focused on the small group discussion. The second phase was dedicated to the design of the "beta" prototype, which focused on how the overall, large group discusses with the help of the small group abstraction among other ideas.

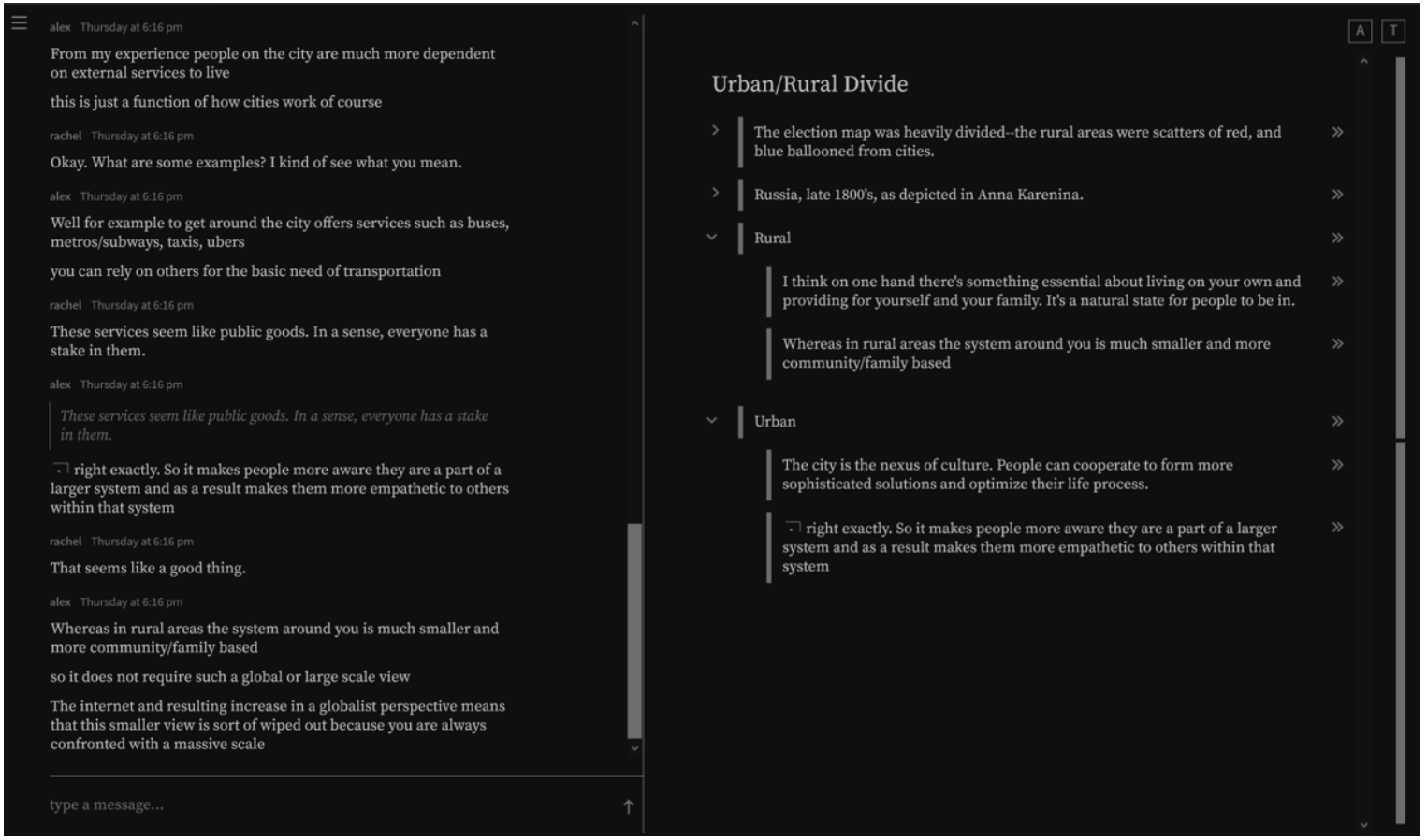


FIGURE 3.1: Example shot of a discussion in the alpha version of Pith.

3.2 Fall: Alpha

Our hope for the alpha version was to allow a small group to embody the ideas of Pith at a smaller scale. In particular, a small group should be able to achieve a productive outcome that could be easily extended to the shared space of the theoretical, larger discussion.

We therefore created an interface with two main parts: the chat for synchronous discussion, and a document to continuously summarize the discussion and would serve as the final product. Even at the scale of a small discussion, a shared space like a document can help the participants keep track of the progress of the discussion.

We imagined that participants would first create a discussion that each of them would join with a nickname, set the topic, discuss the topic in the chat, and move salient points relevant to the topic from the chat to the document to organize and reflect upon. Figure 3.1 shows the final interface for this iteration.

3.2.1 Chat

The chat was composed of posts submitted by the users. Its intention was to allow for mostly synchronous, casual communication between participants. Each post showed the nickname of the author and the time of creation. It had a feature called "transclusion" which is like an advanced version of a reply-to. In composing a post, users could refer to an arbitrary number of other posts instead of one like a standard reply-to feature. When the post was shared in the chat, it displayed a special symbol at the locations in the pith where the user referred to another post. The referred posts were copied above the main post in the order of their reference. We imagined this feature to help users show which ideas they were replying to or were inspired by.

3.2.2 Document

The more novel component of our interface was the document section. Its intent was to help participants summarize their discussion so far and collectively organize their thoughts.

3.2.2.1 The Unit

Unlike most documents, which are intended to be free-flowing, our document encouraged ideas to be presented in units. Each unit had a text, called a "pith". The pith could encapsulate a general theme or topic, an argument, an idea, etc. However, in having individual units, we hoped to encourage people to deliver the particular idea in a succinct manner that others could quickly grasp. In addition, having modular units allowed them to be more easily arranged. Units in the document also had the ability to transclude other units they referenced.

One canonical issue with documents is the problem of concurrency. Several people may want to edit the document at once. If they are allowed to edit the document at the same time, it

is possible their edits will conflict with each other. To address this issue, we had each unit be atomic, that is, only one person was allowed to modify it. Each person would first request a lock on the unit, perform the modification, and return the lock to allow others to get it. Thus, many people could theoretically be modifying the document, but they were guaranteed not to conflict with each other.

3.2.2.2 Hierarchy

The main organization of the document was in a hierarchy. Units formed a tree through parent/children relations. The relation could represent topics with subtopics, subtopics with arguments, or categories with subcategories, subcategories with elements.

Upon the start of discussion, the document contained only one unit, the "root" unit. The users could write the theme of the discussion into the root, from which the rest of their conversation derived from. As units were added, they would naturally be organized in a tree of information, where each unit except the root had exactly one parent and an arbitrary number of children. Units could be moved from one parent to another or from one position on the list of the parent to another position. It was possible that multiple people might want to move the same unit at once. Thus, the only person that was allowed to move the unit was the one who also had edit privilege over the unit.

3.2.2.3 Navigation

Our document was, at its core, an infinitely recursive tree like Workflowy [47]. This created issues in presenting the information. The naive method of presenting the information would be to show the entire tree, where each depth of units was indented one more than the previous depth. However, this could be overwhelming for the user: they would have to scroll left to right for deeply nested sections.

Thus, we decided to restrict how much of the tree a user could see at once and allow them to focus. Each unit had a corresponding page. By default, the page automatically displayed all the children of the unit presented in a list. A user could explore a child in more depth by expanding its own list of children and collapsing it when they were done. If they wanted a more detailed view of the child unit, they could click a button to the right that would allow them to visit its own page.

There were two methods for a user to "return" to where they had been. The first method is through an ancestry line, which had all the units from the root down to the unit of the current page. As the user clicked deeper and deeper into the tree, the ancestry line would grow. The user could click on the top-most entry to return to the parent of the current unit's page, or click on the bottom-most ancestor to return to the root.

The second method was through a timeline, which recorded all the units the user had visited so far. To return to the last visited unit, the user could click on the top-most element of the

timeline. To return to the second-to-last unit visited, the user could click on the second element from the top of the timeline, and so on. Whereas the ancestry line allowed the user to think in a more hierarchical manner, the timeline helped users think of their interaction within the document in a more contextual manner. Through the timeline, they could potentially examine the clusters of their interest over time. For example, the first portion of the timeline might indicate they were generally interested in dog-related topics, the middle portion could indicate they tuned in more to house-related ideas, and the last portion could indicate that they were thinking more about the economy.

For both the ancestry line and timeline, the user could hover over a unit-entry to read the pith of the unit and gain additional context.

3.2.2.4 Coordination

Despite the fact that the document content was shared among the participants, the users could be viewing different sections of it at a given time. One user might be at the page of the root unit, for example, and another might be on the page of a unit nested six layers deep.

There was a row of icons, each representing a user. A user could hover over an icon and see which unit the particular user was on. This feature allowed users to coordinate their locations among each other, such as if they all wanted to talk about the same portion of the document.

3.2.3 Flow of Information

3.2.3.1 Summarization

Perhaps one of the most powerful features was in how the chat and document were connected. During the course of discussion on the chat, any user could decide that a post was noteworthy enough to move directly to the document. This allowed the document to contain the most salient points of the discussion, worthy of preservation and further consideration.

The features of the document itself allowed the information to be further organized. Over the course of the discussion, the document could become the final product of the discussion.

3.2.3.2 Linking

Both the chat and the document had a keyword search capability, which is common among other chat-like systems, such as Messenger [20], Slack [21], and Discord [22]. However, the intent of search was focused on how to link a new unit to other units. A user could perform a search as they were composing a post or the pith of a unit on the document. The search allowed them to find posts related to the query of interest and then reference them in the current post. Thus, the combined searching/linking capability allowed ideas to feel more accessible, encouraging users to consider the viewpoints of those that came before them.

3.2.3.3 Small to Large

We designed the alpha so that it could be extensible. We imagined the alpha to form the basis for what a small discussion would be like. In addition, we imagined that there was some way by which the small discussion, embodied in the alpha, could interact with the overall large discussion.

Our general conception of this was that later on, users should be able to view the representation of the large discussion and be able to add to it near the conclusion of their small discussion. We imagined that the representation of the large discussion would also be a hierarchical document that users in the small discussion could examine in a similar fashion to how they viewed their local document–with an ancestry tree, a timeline, and search. We imagined that users could then add sub-trees of their local discussion document-tree to some unit-node in the large discussion document-tree.

3.2.4 Other Ideas

We had a number of other ideas that we considered or even tried but ultimately did not include. However, these ideas often helped us arrive at the above design for the alpha. A summary of these ideas is presented in Figure 3.2.

3.2.4.1 Backlinks

Our system allowed units to reference other units. However, we were interested in the idea of allowing units to backlink other units, similar to how one can use backlinks in Roam Research [51] and Notion [52]. That is, if unit A linked to unit B, unit B would have a backlink to unit A. Our hope with this idea was to allow greater interconnectivity of ideas. If unit B had a backlink to unit A, then users interested in unit B could be interested in its context within other units, such as unit A. They could have greater exposure to how a certain idea was used. The link/backlink structure allows an idea to be further replicated. The link/backlink structure is somewhat similar to a thread of posts, in that in its extended form, it represents inspiration over time. Except, instead of authors replying to each other, ideas embodied in posts take inspiration from each other.

One add-on idea was to allow users to view the network of a unit, to see all that had linked and backlinked to it, to see all units that could have been inspired by it and that it took inspiration from. Unfortunately, we did not have time to implement the add-on or the backlink feature during this phase.

3.2.4.2 Tags on Posts

One of our first ideas was to have tags on posts. We were loosely inspired by Twitter's hashtag feature [15], which allows users to quickly grasp the context of a post. We created a simple

chat where users could add and remove tags from any post, even those they didn't author. We enlisted a group of people to informally test the chat. We quickly realized the power of tags to detract user's attention from the main discussion. Instead, users would have mini-discussions on a post, as if trying to continue a thread. The tags were also often used as ways to express emotions about a post, but to do so in a way that was often annotated. This was undesirable, because it encouraged people to make snap judgments without critically considering why they felt that way. If users had instead formulated their reaction in a post, addressing the whole group instead of the post or the author, they would be encouraged to offer a more thought-out explanation for their opinion. Thus, in our alpha prototype, we no longer had tags, instead focusing on the unit as the atom of information.

3.2.4.3 Flairs

Another idea we had was to allow users to flair themselves, by offering short phrases that indicate some aspect of their identity, as is done in Reddit [16]. For example, a user could label themselves as "expert", "hoping to learn", "citizen of Pittsburgh", etc. Flairs can help users get a sense of who they are talking to and what each person's potential contribution might be. We eventually didn't include this more for the sake of time. However, it may be worth further understanding how this feature could be used by people if it is integrated into our system.

3.2.4.4 Saving

Some chat systems, like Messenger [20] and Slack [21], allow users to save a post for their own record. We considered experimenting with this idea but eventually ended up not doing so due to time. Instead, we were more focused on how a group could share ideas, similar to pinning a post in Slack or Discord. Through a "pin-like" feature, people could work together to decide which posts were worth considering further in their summary.

3.2.4.5 Cursor

We were concerned about how people would coordinate what they saw on the document as they were discussing. Though we eventually settled on having each person browse through the document through pages on their own and for them to use user icons to determine where others were, we started with a more fine-grained approach. Our original idea was to use a special kind of cursor. Each user would have a corresponding cursor indicating not only which unit page they were viewing but where on the corresponding list of sub-units they were viewing. They would use the cursor to add new posts exactly where their cursor was on the current page. However, this method was cumbersome to program correctly and worse, would probably be unintuitive to most users. Thus we simplified the design.

3.2.4.6 Summary Limits

A summary is probably not useful if users move most of their discussion, most of the chat posts, to it. One of our ideas was to explicitly encourage people to keep their summary short. For example, by limiting the number of characters for a unit in the document or limiting the number of units in the document itself. The limit could be hard-coded into the system or set by the users of the system. However, we eventually decided such a constraint was likely not necessary, that people could organically determine how much space was needed for their current application.

3.2.4.7 Topic Phases

We noticed that in the course of a discussion, the general topic could shift such that a topic visited later could have no relation to a topic from before. We proposed the idea of phases, that upon each phase, each topic, a discussion group would have a new document they could use to summarize. Users could move between the phase-documents if they happened to wind back to an old topic during the course of their natural discussion. However, with the tree structure of the document, this structure was not necessary. Users could create topic-units on the root unit and and delve into the sub-tree of each topic-unit.

3.2.4.8 Chat Reduction

In a natural conversation, a group typically remembers what was said up to a point in the past. We wanted to mimic this in our system by allowing only the last N (i.e. $N = 50$) posts to be preserved, such that upon the addition of a post, the oldest preserved post would be forgotten. We were hoping this mechanism would encourage people to save posts into the document they thought was meaningful, lest it be lost.

3.2.4.9 Discussion Expiration

We were originally imagined small discussions to be of finite duration so that the participants would be able to pace themselves in accomplishing some goal in the course of the discussion. Thus, we implemented a feature where people could set the lifetime of a discussion, such that at the end of the lifetime, it would expire.

3.2.4.10 Summary Export

Because the document is meant to act as a summary, it was likely that people might want to save the summary to present through other means. We thus experimented with several export options, such as to export the document in a PDF, HTML, or simple-text format. This could allow for greater interoperability with other utilities.

3.2.4.11 Turn-Taking

One of our concerns was how to encourage participants to "listen" to each other through the design of the platform. Because Pith is a text-based platform, "listening" would be akin to carefully reading a post and reflecting on it. A common mechanism used in real life and in audiovisual platforms like Zoom [25] is for people to take turns. People can raise their hand and be called on to take their turn. However, hand-raising requires a moderator to decide who goes next. A natural way to moderate who goes next would be to use a queue. People could place themselves at the end of the queue, and whoever was in front would go next by submitting their post for everyone to consider for some period of time. We didn't implement this in the alpha prototype, but it could be worth implementing for future prototypes.

3.2.4.12 Acknowledgment

Another way we thought to encourage listening, or careful reading of posts in this case, was to use an acknowledgment mechanism. A user would check off each post in the chat they had read. Only if they had read all the posts generated so far could they create a new post. The natural problem with this approach is that people might game the system by checking on each post, even if they hadn't read it. However, the relative simplicity of this idea was appealing to us. It could be worth experimenting if a group can self-moderate so that this mechanism or similar mechanisms are used as intended.

3.2.4.13 Cultures

Moderation is a serious problem on many discussion platforms, such as Reddit [16]. Moderation is used for a group to enforce a certain culture, a set of guidelines and rules for users to follow while participating. A common rule is often to be polite to other participants. Some mechanisms to enforce rules are the ability for users to flag posts as inappropriate or for users to vote to ban someone from the group.

One of our ideas was to eventually create a myriad of various moderation tools and allow each community to create their own culture by choosing their own set of rules and mechanisms for enforcing these rules. Pith was imagined to be an open-source technology that theoretically anyone could spin off a server for and use as they desire, similar to the concept proposed by Mastodon. Thus each community would be self-contained and self-moderated. The default structure of each community could be a direct democracy with a constitution of rules and enforcing mechanisms. The community could decide that having dedicated enforcers is the most effective way to moderate, and thus write this in their constitution. This is in contrast to Reddit's moderation structure, where each group, by default, has several designated moderators [16].

The culture concept could extend beyond simply the rules of conduct to encompass how a community desires the overall discussion to be organized. For example, one feature could be for

a community to decide that all small discussions should be timed a certain duration so that the overall discussion could progress in a more timely manner.

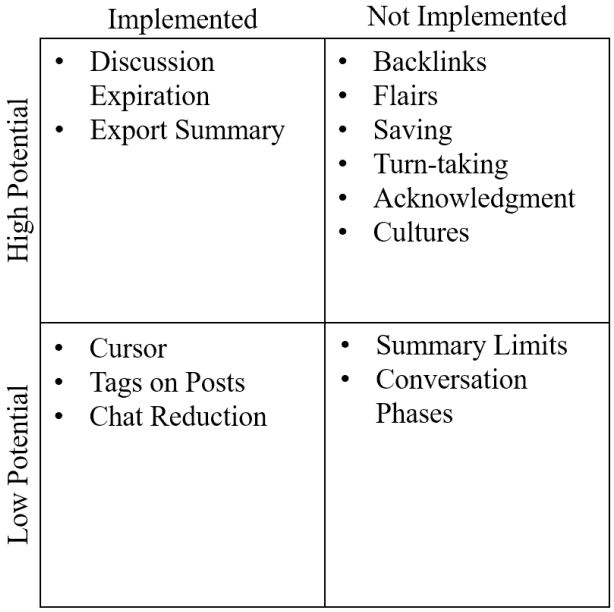


FIGURE 3.2: Categorization of alpha features into implemented and not implemented, and whether we believed it had potential or not.

3.2.5 Reflection on Alpha Design

We used the prototype ourselves to understand its advantages and disadvantages. Our main observation was that despite our aim to create a simple, organic system, the system was too complicated. The document structure was too rigid. Because units of ideas were arranged in a tree, it was more difficult to see how ideas in separate branches could be related to each other and more difficult to draw connections between them.

We had previously anticipated this problem, which was why we proposed link/backlinks to overcome some of the rigidity. However, we suspected that even if we had implemented links and backlinks, the problem would persist. The issue seemed to be that the basis of organization was hierarchical. Even using the ancestry tree and timeline features, the document was difficult to navigate. However, as people are first discussing ideas, their thoughts are more loose, not as well organized. It could be of greater benefit to allow them to be more flexible in organizing their information, in seeing relationships between ideas as they are discussing.

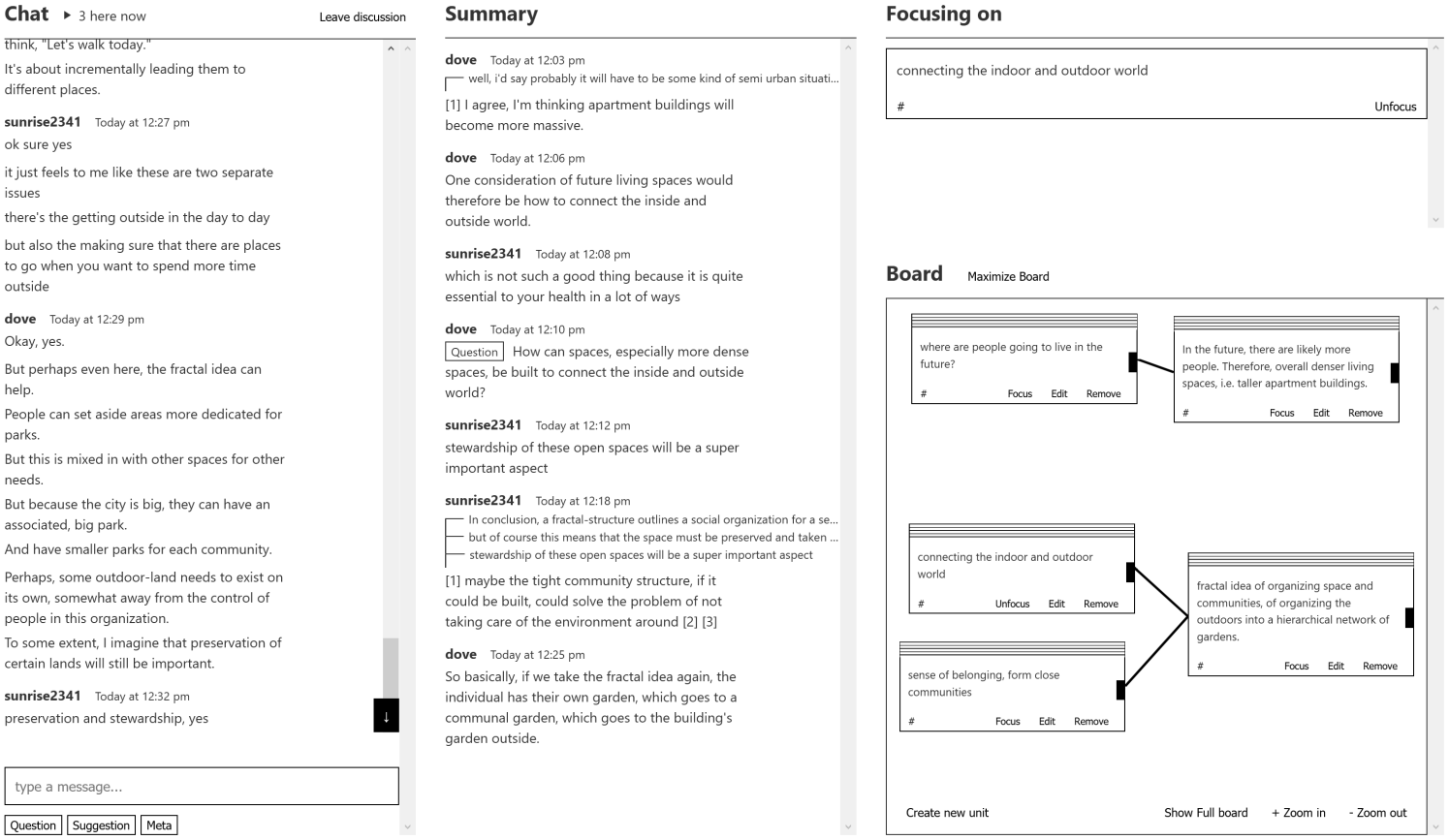


FIGURE 3.3: Example shot of a discussion in the beta version of Pith.

3.3 Spring: Beta

Our main goal during the beta phase was to prototype a system that allowed a large group to discuss, and from which small discussions could originate and contribute to the overall discussion. Our beta deviated from the design of the alpha. The hierarchy was no longer the default structure to organize information. Instead the default structure was the network.

The group discussion was also a document of sorts, but instead of being a recursive list, it was a 2D "map" called the board with linked units. The board served a similar role to the document discussed before, in that it represented the state of the overall discussion at the current point in time.

Small discussions would originate through the current state of the overall discussion. To explain how this works, let us consider an analogy. Let us imagine that the board is like a physical map, such that each unit on it is like an island. A small discussion would be like a boat. Just as a boat can be created on any island, a small discussion could be created from any initial unit on the board. Just as boats can travel from one island to another during the course of their journey, a small discussion could move from one unit to another as the focus of the conversation changed. For example, a small discussion may have originate on the topic-unit of "mitigating anthropogenic climate change", then shifted to the topic-unit of "renewable energy technology", then to "U.S. rural economy".

3.3.1 Board

The board served as the representation of the state of the overall discussion. It was a 2D map to organize a network of information. Units of information could be added directly, such as to decide on topics of discussion. Over the course of the overall discussion and the further spawning of small discussions, more units could be created as more ideas and topics came up. People could move units around on the map, such as to group related units together and to consider possible relations between then. Links more formally encoded relations between units and helped draw people's attention to the found connection between the ideas/topics/themes embodied by the units.

3.3.1.1 Units and Links

The board started with a given size. However, if people wanted to add more units than there was space available, they could pan across the board to reach free space and zoom in and out. Units could be added, edited, or removed by anyone. Anyone could also create unidirectional links between units with a description of what the link meant. For example, a unit that said "renewable energy technology" could have a link to "U.S. rural economy" that said "improve through job creation". Links could also be removed.

3.3.1.2 Context

A user could explore an individual unit in more depth by clicking on it. By doing this, they could see all the discussions under it, as well as an ego network of the unit, a network of all the other units that linked directly to or from the given unit.

3.3.1.3 Beams

We desired to create mechanisms for group meta-cognition, the ability for a group to reflect on themselves or in this case, on their discussion. Beams was the fist mechanism by which to encourage this.

People could create a special kind of unit on the board, called a beam unit, that was highlighted and thus differentiated from other units. The purpose of the beam unit was to suggest some high-level change in the overall discussion. For example, what topics should participants focus on? How should the board be reorganized, especially if there are too many units? The beam unit had two fields. The first was the pith field, which every unit has, and was where they would put the suggestion. The second was the rationale field, where they could put why they thought their suggestion was meaningful or should be followed. Just like any other unit, people could create discussions on a beam unit and link them to other units or beam units.

3.3.1.4 Reflection Periods

Our second idea to encourage group meta-cognition was to allow users to create reflection periods, a feature intended to work in tandem with beams and that we plan to implement in the future.

Most of the time, people would likely default to generating new units and links, to adding their own or their small discussion group's ideas. However, if most effort was dedicated to generating ideas on the board, the board could become unwieldy and disorganized. The focus of the overall discussion would become unclear.

Through the reflection period feature, people could create reflection events, possibly periodic, dedicated for reflection of the current state of the overall discussion, as embodied by the board. We imagined they would spend the time during these periods to address the state of the discussion itself. For example, what topics need to be addressed given what has already been brought up? How can the board be organized to encourage participants to focus on the most essential topics? How can the board be organized to be easier to navigate? Given that the beam units should embody many of the possible suggestions, we imagined that during the reflection period, people would focus on joining a small discussion on one of these beam units. They would examine the board and use additional beam units to coordinate meta-cognitive ideas between the small discussions. Then they would coordinate to enact their changes.

We believed encouraging people to consider having dedicated events to meta-cognition would enable them to better manage the overall course of their discussion.

3.3.2 (Small) Discussion

A small discussion could be created by any individual on any unit on the board. Other individuals hoping to join a small discussion could click on whatever units on the board interested them, then choose a discussion listed in the unit. Once an individual joined a small discussion, they would see both the board and the small discussion interface. The interface contains three main components: a chat section, a summary section, and a focused section.

3.3.2.1 Chat

The chat's purpose was to allow a natural conversation between participants. It was mostly similar to the alpha version in that it had transclusion and a search capability to find posts to transclude.

The main change to the chat was the inclusion of mechanisms to encourage active listening.

Through our experiments, we developed several methods to potentially address this.

Typing Indicators When at least one person was typing, the chat would indicate to everyone else the number of people that were typing, but not display who. Our hope was that through this particular version of typing indicators, people would self-moderate in deciding whether or not to type their post. If they didn't type a post or weren't "speaking", we anticipated they would instead be attentive to posts that were being generated and read them more carefully. Ideally, more people would decide to wait and "listen" versus type and "speak" at a time.

Post Flairs Another method we experimented with was the use of flairs on post. If users focused on what they added to and removed from the summary during the course of their discussion, posts with suggestions to modify the summary would probably be of greater interest than regular posts. We hypothesized that having a "Suggestion" flair on a post could therefore help draw people's attention to the worth of a post. We created three flairs: the "Suggestion" flair, the "Question" flair, and the "Meta" flair. The "Question" flair could be used to highlight a post that asked a question to the group for them to answer, or a question the asker thought needed a response. The "Meta" flair was intended to be used in cases where people suggested high-level action, such as changing the discussion to focus on topic B rather than topic A.

A flaired post would also be automatically added to the summary section. The first reason for doing this is more practical: to encourage consideration of the post, to promote active listening. The second reason is more conceptual: the flaired post would be a highlight of the discussion, and to put it on the summary would help people remember it and deliberate on it soon.

3.3.2.2 Summary

The discussion had a summary section with a similar purpose to the document in the alpha prototype. Participants could move posts of interest from the chat to the summary to save. As

they pinned or unpinned a unit from the summary, a special unit that we call a "notice unit" would appear in the chat broadcasting to the group about the action and who performed it.

Unlike the alpha's document, the summary was not a hierarchical, recursive list but a single list. Posts could only be added in order and not moved around. Thus the summary was far simpler than before.

A subtle design choice we made was to have the chat and the summary be the same size. Both took up the height of the screen and had the same width. Our intention was to encourage people to think of the summary as a critical component of their discussion, to use it to condense their conversation to its most salient points.

Posts could also be removed from the summary. Over the course of discussion, a summary post could no longer be considered essential, and as such could be removed.

We implemented a feature to directly "publish" a summary post to the board. This allowed an idea to be moved from the chat, reflected upon through its inclusion in the summary, and added as a contribution to the overall discussion by sending it to the board.

With this feature, it was also possible to publish a post directly in the chat for ease. If the overall group was so small they could all fit in a small discussion, they might decide to use the board to organize their ideas rather than using the summary section as an intermediary.

3.3.2.3 Focused

If the small discussion was a boat, the focused section was a like a set of tacks indicating which islands, or units, the small discussion was on. Its intention was to allow the group to focus their conversation around a set of ideas. Multiple units could be in the focused section at once. At the start, only the unit the small discussion had been generated under would appear in this section.

We used notice units to broadcast actions taken on the summary. We also plan to extend the appearance of notice units when small discussion members interact with the focusing section by focusing or unfocusing units. In addition, notice units can be used to indicate changes made by small discussion members to the board, such as by adding units there.

3.3.3 Other Ideas

The following highlights ideas we experimented with in addition to our main design. A summary of them is presented in Figure 3.4.

3.3.3.1 Originating a Small Discussion

Our current methodology to group people into small discussion groups was to have people distribute themselves among small discussions partly based on interest and partly based on chance.

One of our concerns was that some people might want to have an unfocused discussion simply to explore different ideas. It is possible to achieve that on the current design by creating a unit on the board with a blank pith, for example, and having such small discussions be created there. But it may be desirable to create a more explicit mechanism to have an unfocused discussion.

Another possibility is that people may decide they want to talk to specific users for whatever reason. Then they may want to create a discussion with pre-chosen participants. For now we have not implemented this feature, but it may be worth considering if it seems to be necessary.

3.3.3.2 Amendments

One of our concerns was how to mediate edits to a unit, the removal of a unit, or removal of a link on the board. It is possible that someone could perform one of these actions unintentionally or maliciously. We thus considered a framework where each operation could only be suggested in an amendment, but could not be automatically performed on the board. Its approval would then be decided by some sort of voting process.

3.3.3.3 Explicit Stakeholders

In continuation of the issue about managing edit rights, we wondered what mechanism should determine who should have the ability to decide what amendments should be approved? Who should have the general responsibility of overlooking a unit or set of units? We considered that one way to address this was to create the notion of explicit stakeholders. By default, a created unit would have no stakeholders. That is, no one would care very much if it was edited or removed. Thus any edit or removal on it could be done automatically. However, should an individual decide they cared about the maintenance of some unit, they could indicate they were a stakeholder of the unit. If there was even one stakeholder on a unit, any edits or removals relating to the unit would have to be approved by them first. If there was a group of stakeholders interested in a unit, they would have to decide together whether a potential edit or removal, embodied in an amendment, should be allowed.

Originally, the idea of amendments and stakeholders was used to address the approval issue on a document organized more as a hierarchy rather than as the network structure we incorporated. Thus, certain operations, such as moving a unit from one parent to another, were especially complicated to consider.

We did not use these concepts in our beta prototype. Instead, every user was allowed to edit and remove a unit or remove a link for simplicity of design. However, it may be worth to consider these ideas further, especially after observing real users.

3.3.3.4 Discussion Phases

A precursor to the reflection period was the idea of using discussion phases. A discussion can be thought of as an iterative design process to formulate some outcome, such as an action. The iterative design process has two main phases: generative and reflective. In the generative phase, participants focus on brainstorming, or generating, ideas. In the reflective phase, participants reflect on refining the set of ideas, such as by having more focused discussions on whether each is worth considering further. In application to the Pith-based discussions, we imagined a similar process. A group could decide to organize their overall discussion into phases of generation and then reflection. During generation, they would add more ideas to the board, and during the reflection they would primarily edit and remove the existing ideas. Thus, along with this idea, we considered that the operations of edit or remove should only be open during the reflection phase.

3.3.3.5 Attention Metrics

Similar to Klein (2012) [39], we considered several metrics to draw people's attention to particular units on the map. In particular, we considered priority, explorability, and activity. Priority of a unit would be derived from a dynamic, user-based weighting of how important it is and requires discussion. Explorability is calculated from a dynamic, user-based weighting of how much potential people believe an idea encapsulated in a unit has. Activity indicates how much activity there currently is surrounding the unit, such as the amount of discussion or the number of linked units. A unit with high explorability or priority but low activity could potentially encourage more users to join the discussion.

However, there are several issues with this proposed system. The activity metric have the effect of drawing people to popular units rather than to under-served units. In addition, we considered that explorability and priority as a metric may be unhelpful, because they do not indicate why the unit is highly explorable or should have high priority. Thus, after further reflection we decided to start with beam units, which could allow people to justify to others why they believed a unit should be prioritized or explored.

3.3.3.6 Hyperbolic View

The board provided a flat view of all the units, such that all units are the same size. However, should the board grow in size, a user may desire to see the full scope of the board and simultaneously focus on a particular region of interest. The ORA tool created by the Center for Computational Analysis of Social and Organizational Systems (CASOS) group here at Carnegie Mellon University has a feature called the hyperbolic view [59]. The hyperbolic view allows a certain section of a network laid out in 2D space to be blown up in the visualization, such that the surrounding area is visible but minimized. This feature allows one to see both the macroscopic structure of the network and microscopic structure of the region of interest. This

feature could be of use to help Pith groups with a large network of units on their board analyze it more effectively.

3.3.3.7 Unit Flairs

As discussed before, units contain the atom of information on the platform. However, the pith of the unit may describe a topic, an idea to consider, an argument or a support, a piece of evidence, a fact, or an anecdote. It may help people categorize what kind of information each unit is to understand its intention in the overall discussion. Thus, we proposed flairs for units to do just this. The flairs could be preset or could be created in real-time. If they are created in real-time, there is the possibility they may also be treated as tags, which could have its own set of pitfalls as described previously.

We already implement a version of this on the small discussion, but it would worth considering the different flairs that would be most useful to people and to implement this for use on the board as well.

3.3.3.8 Polls

Polls are useful to make decisions quickly. We considered allowing a small discussion to create them, or even allowing it to be a special unit on the board to allow larger votes to be made. It would be interesting to study when users feel compelled to use a poll to determine if the voting hinders or stimulates further productive discussion.

3.3.3.9 Sets

Similar to the concept of tags was our proposed concept of sets. Through set-relations, units could have multiple children and multiple parents. Thus, sets was meant as a way to formulate a more flexible hierarchy among ideas.

The unit itself acts as a set that groups its children. For example, there may be a "renewable energy" unit with "solar power", "nuclear power", and "wind power" as its children.

On the board, one could click on a parent-unit, which acts as the set descriptor, to "collapse" the set and hide all children units, to "expand" the set and re-show the children units, or to "coalesce" the children units by temporarily bringing them closer to the parent-unit on the board so they can be visualized close together in space.

3.3.3.10 Portals

One of our most intriguing ideas was the concept of portals: special units that allowed two boards to be linked together.

Generally, it would be desirable to keep the board focused on the current discussion on hand and to be as compact as possible. Conversation around a set of units, likely around a similar theme, could grow so intense that it detracted from the flow of the overall discussion. Thus, these offshoot units and their small discussion on the current board could become the progenitor of discussion on a new board. Or, if an overall discussion had clearly separate but related subtopics of discussion, a new board could be made for each of these subtopics, accessible on the main board. Because there would still be some relation between the discussion on the original board and the offshoot board, it would be useful to connect them together.

A portal allows the original board, call it board A, to have a link to the offshoot board, call it board B. A portal is still a unit, and thus can be connected to other units and have discussions under it. By clicking on the link, users on board A can go to board B and see the state of the discussion there. In addition, should a portal be created from A to B, B would have a back-portal to A. If users were to click on the back-portal, they would be able to see the context of the portal-unit on board A, what other units it is linked to, the discussions around it, etc. to get a sense for the context of board B's discussion space within board A's.

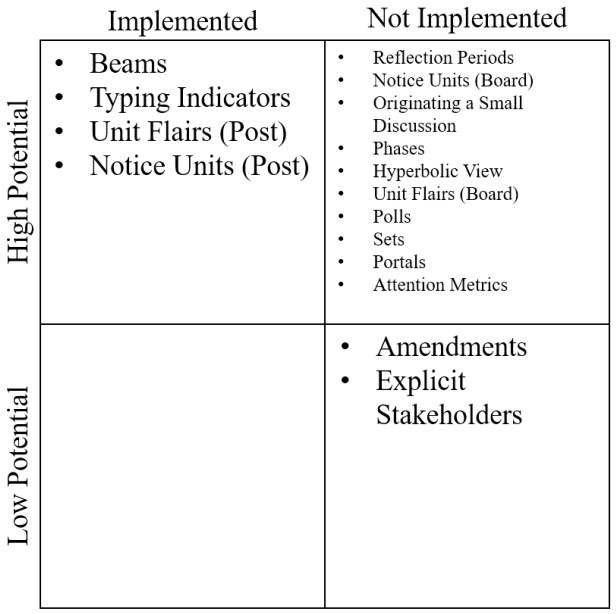


FIGURE 3.4: Categorization of beta features into implemented and not implemented, and whether we believed it had potential or not.

Chapter 4

System

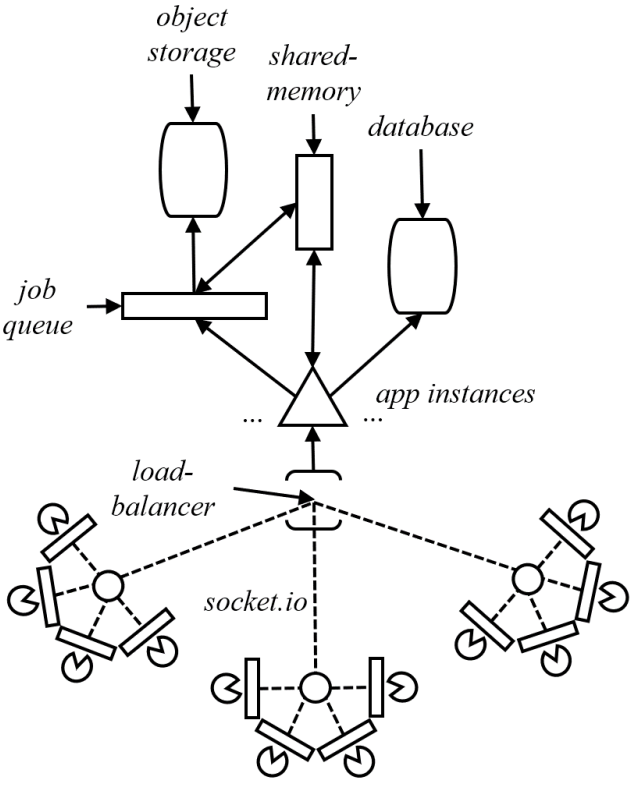


FIGURE 4.1: Schematic of system, primarily showing backend components.

4.1 High-Level Hardware

Pith was a website, with a frontend to present the interface for a user and the backend to handle the logic of the discussion. Figure 4.1 shows the key components in the backend, which I primarily implemented. My partner, Christian Broms, primarily implemented the frontend.

There was at least one instance of the app server, which served discussion and board functionality to clients. The client and server communicated back and forth through socket.io [63]. Socket.io has a concept of rooms, which is a defined group of clients. If a response is transmitted to a room, it is transmitted to every client in the group. We used this abstraction to mediate communication between users in the same small discussion or users on the same board.

For heavier loads, multiple app servers could be used, and an HAProxy load balancer would mediate which client was served by which server such that all their requests in the session were served by that particular server [64]. A Redis unit allowed the multiple servers to have a shared memory space to communicate, for example, how to properly emit data to a room [65].

Data about the small discussions or boards were stored within a MongoDB database [66] and retrieved or updated by the servers primarily through client requests.

The app servers would put more time-consuming jobs onto an arq job queue. Jobs may have needed to store data, such as an export file that could later be retrieved, on the object storage.

The arq job queue, Redis shared memory, MongoDB database, load balancer, and app server each ran in their own Docker container [67]. A MongoDB cloud cluster could also be used in place of the MongoDB docker container. An additional Docker container would serve the frontend code and static pages to clients.

The frontend was built in Javascript. During the fall for the alpha prototype, React was used as the framework [68] with Redux for state management [69]. We switched over to using Svelte and Sapper to build the frontend because of its greater simplicity and flexibility for our purposes [70, 71].

Refer to Appendix A for further citations of tools used.

4.2 High-Level Software

Figure 4.2 summarizes how a request from the client was processed on the backend.

Through socket.io, we generated a user session for each client. Upon joining a board or small discussion, the corresponding name of the room was stored in the user session. Upon leaving, the name was removed. The client could only be in at most one board at a time and at most one small discussion at a time. When the client made a request that was meant to be transmitted to the small discussion, the server used the room in the client's user session to broadcast the request to the members of that room. For example, if a client desired to post a message in the

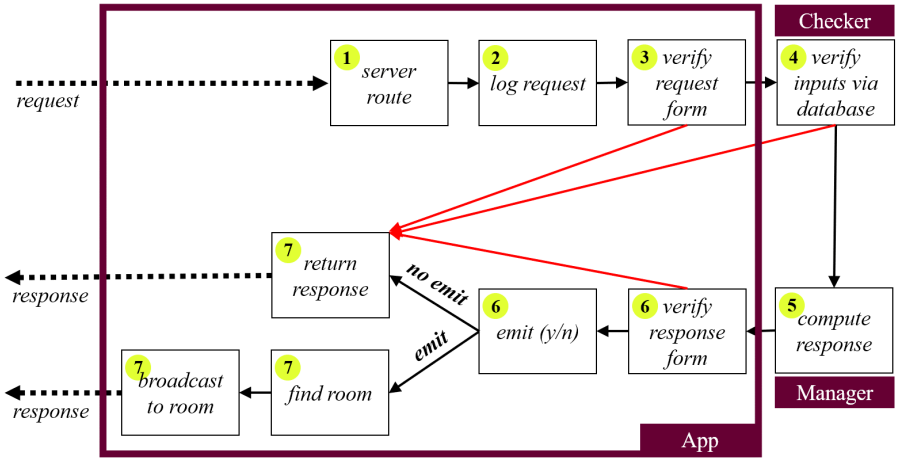


FIGURE 4.2: Steps of processing a request from a client on the backend. Red lines indicate the occurrence of error. Maroon indicates which component handled the step.

chat of a small discussion, the server would find out which small discussion the client was in and send the response containing the information about the new post to everyone in the small discussion.

A client's request was in JSON and passed along a route, which specified what kind of operation the request was; for example, the operation to post to a chat. The request was logged for maintenance and debugging purposes.

For robustness, we had several levels of verification. The JSON request was validated using jsonschema to ensure it had the desired inputs for the operation [72]. If the validation failed, a request error would be returned. However, the jsonschema validator only checked the form of the inputs, such as whether there was a argument named "text" of type `string` in the request. Some inputs within the request needed to be checked against the database by the Checker. For example, one input could be the user ID of the author. A check would be to determine if the user ID existed in the database of users. If the check failed, an error was returned to the client. Once all inputs were determined to be valid, the Manager would process the request to generate a response. The response, also in the form of a JSON, would be validated with jsonschema as well. If the validation failed, a server error would be returned. Should the validation succeed, the response was either returned to the original client or emitted (transmitted) to the room of the client, based on the operation type.

Because the alpha and beta prototype had different designs, they had different implementations. In the following sections, we go over the details of how different aspects of the design were abstracted, the operations clients could perform with the abstractions, and notable issues in our system we discovered or resolved.

4.3 Fall: Alpha

4.3.1 Schema

We represented several key types of objects on the backend through `mongoengine` schemas, which specifics the attributes of each object. The key abstractions we used were for the small discussion itself, the unit, and the user.

The approach taken to design each schema was "what does each type of object need to know relevant to itself?" This dictated what attributes each type of object had. Because of this, some relationships were encoded multiple times. For example, the relation of User X has lock on Unit Y would be encoded in User X and Unit Y.

4.3.1.1 Discussion

Discussion objects comprised the `discussions` collection in the database.

1. `id` - the unique `string` representation of the Discussion in the database.
2. `document` - the root Unit of the document associated with this Discussion. Acted as the root of a tree structure.
3. `chat` - the list of Units in the chat associated with this Discussion.
4. `users` - the list of Users in this Discussion.

4.3.1.2 Unit

Unit objects comprised the `units` collection in the database.

The Unit was abstracted to be independent from the Discussion rather than be strictly associated with one Discussion. This was mostly to anticipate the possibility that Units from one Discussion could be referred to from a Unit in another Discussion to encourage further propagation of ideas. The Unit also stored which Units it was connected to in the hierarchy of the document, its parent and children. It also recorded who had the "lock" for editing it or moving it, as only one User could be doing either at a time.

1. `id` - the unique `string` representation of the unit in the database.
2. `pith` - the text within the Unit.
3. `discussion` - which Discussion the Unit originates from.
4. `in_chat` - whether the Unit was created in the chat (`true`) or in the document (`false`).
5. `children` - the ordered list of children Units of the Unit.
6. `author` - only applicable if the Unit is in the chat; the User who wrote the Unit.
7. `created_at` - only applicable if the Unit is in the chat; when the Unit was created.
8. `parent` - the Unit that is the parent of the Unit.
9. `edit_privilege` - the User that is currently editing the Unit.
10. `position_privilege` - the User that is currently moving the Unit.

4.3.1.3 User

Formally, the User was embedded within the Discussion schema; that is, the User could only be associated with one Discussion. If a person wanted to be part of multiple discussions, they would need to create a persona for each discussion.

We stored the state of the user on the discussion. A user could be in the discussion but not online (inactive) or online (active). This information was needed to compute, for example, how many people were actually active in the discussion room. We also stored what unit page they were visiting so we could reload it when they came back online.

We also recorded whether the User had any Unit locks. It was important that these locks be released when the User left so that others had the ability to edit or move the associated Units.

1. `id` - the unique `string` representation of the User in the Discussion.
2. `viewed_unit` - current unit page User is on.
3. `name` - name of the User as seen by everyone in the Discussion.
4. `active` - whether the user is online or not.
5. `timeline` - the list of Units the User has viewed. Used to propagate their timeline on the interface upon start.
6. `editing` - what Unit, if any, the User is currently editing.
7. `moving` - what Unit, if any, the User is currently moving.

4.3.2 Operations

The following specific operations were implemented in accordance with the alpha prototype design.

A client needed a link to access the discussion. The link encoded the ID of the discussion.

1. `create()` - creates a discussion.
2. `test_connect(discussion_id)` - tests that the given ID, most likely in the link the client received, is of an actual Discussion. If not, the frontend should render a 404 page that the link is invalid.
3. `create_user(discussion_id, nickname)` - creates a user with the given nickname.
4. `join(discussion_id, user_id)` - lets the user join the discussion. Returns data the user needs to load the discussion information, such as the unit page the user was last on, their saved timeline, and the chat history. Emits to the other users in the discussion room that the current user has joined.
5. `leave(discussion_id, user_id)` - lets the user leave the discussion. Emits to the other users the current user has left.
6. `load_unit_page(discussion_id, user_id, unit_id)` - returns to the user the information about the given unit so the client can load the unit page. Also emits to the other users that the current user is viewing the given unit.

7. `get_ancestors(discussion_id, unit_id)` - gets the ancestors of the given unit. Often used to build the ancestor navigation mechanism for the client.
8. `get_unit_content(discussion_id, unit_id)` - returns the pith for the unit, useful for generating information when a person hovers over an icon of a unit for more details.
9. `post(discussion_id, user_id, pith)` - creates a new post authored by the user with the given pith. Emits post to everyone in the discussion.
10. `search(discussion_id, query)` - performs a search over the units with the given query. Useful for transclusions.
11. `send_to_doc(discussion_id, user_id, unit_id)` - copies the given unit, which should be in the chat, to the document under the unit the user is currently viewing the page of. Emits to everyone in the discussion about the addition.
12. `add_unit(discussion_id, pith, parent, position)` - adds unit with a given pith to the particular position under the parent within the document tree. Emits to everyone in the discussion about the change.
13. `select_unit(discussion_id, user_id, unit_id)` - lets the user attempt to take the lock for a unit to prepare to move it. Emits to everyone in the discussion so they can no longer try to get the lock and move the unit.
14. `deselect_unit(discussion_id, user_id, unit_id)` - lets the user release the lock to move a unit. Emits to everyone in the discussion so they have the option to try to get the lock and move the unit.
15. `move_units(discussion_id, user_id, units, parent, position)` - lets the user, who should have the move-lock, move the group of units under the parent at the given position. Emits to everyone in the discussion about the change.
16. `request_to_edit(discussion_id, user_id, unit_id)` - lets the user try to take the lock for a unit to prepare to edit it or remove it. Emits to everyone in the discussion so they can no longer try to get the lock and edit the unit.
17. `deedit_unit(discussion_id, user_id, unit_id)` - lets the user release the lock to edit a unit. Emits to everyone in the discussion about the change so they can reclaim the right to try to get the lock and edit or remove the unit.
18. `edit_unit(discussion_id, user_id, unit_id, pith)` - lets the user edit the unit with the new pith. Emits to everyone in the discussion about the change.

4.3.3 Issues

4.3.3.1 Concurrency

The proposed mechanism for managing concurrency of editing and moving units, by keeping locks mediated by the server, is in practice too slow. There is some network delay in first acquiring the lock and then making the change. When we tried the system ourselves, this delay was very noticeable.

4.3.3.2 Test Suites

We originally wrote formal test suites to verify assumptions about the behavior of the backend operations. However, this grew difficult to maintain. The codebase itself often changed as new issues in the design or the implementation were continuously uncovered. Thus, to update the test suites in accordance with these changes often doubled the amount of effort needed as compared to not updating the test suites.

4.4 Spring: Beta

4.4.1 Schema

The approach taken to design the schema was "what does each type of object need to know about itself *and* how can we encode this information only once throughout the whole database?" The main desire to having only one representation of some information, such as relationships between two objects, was to prevent conflicts between different representations and the possibility for more bugs to arise.

We designed each schema to be more compact than before. Each MongoDB object can hold a limited number of objects within itself, such as a limited number of entries in a list attribute. Therefore, it would not be scalable if a Board object should hold all the User objects associated with it, as the number of users on a board drastically increases. Furthermore, a minor update to an object in the database would overwrite the entire object.

The main way we achieved condensed schemes was to avoid having them encode the storage of a type of relationship for which the object was expected to have arbitrarily many of. Instead, the relationship would be represented by a new type of object and stored within the collection associated with that type of relationship. For example, consider the link type of relationship between two units. In the alpha implementation, the link between two units was stored within both corresponding Unit objects. However, in the beta implementation, the link between the units is represented as a Link object within the `links` collection. If we consider the user-in-board type of relationship, we would similarly want to avoid storing all the users of the board within the Board. Instead, in the beta implementation, only the User would encode this relationship, as a user should only be associated with one board.

We anticipated that a discussion should not have much data in the chat, summary, or focused sections. Thus, the Discussion objects in our beta implementation still contained the lists of all the Units that were in each of these three sections for convenience. Further testing could reveal whether this is a reasonable assumption to make.

While MongoDB is not so vertically scalable, in the sense that each object can not be large, it is extremely horizontally scalable. We took advantage of this to design our collection organization. All the objects of some type, such as Discussion, across all the boards, would be stored in the same collection. However, to help identify a Discussion object from this mega-collection, the

object would store the ID of the Board it was associated with as well as a short ID, a unique ID among the Discussions associated with that Board. The unique ID of the Discussion object was then the concatenation of the Board ID and its short ID. This formulation of unique IDs was done in anticipation to be used for ranged sharding should the amount of data grow and more servers be needed. Then objects in the same collection associated with the same Board would more likely be stored on the same server, closer together.

4.4.1.1 Board

Board objects comprised the `boards` collection in the database.

1. `id` - the unique `string` representation of the Board in the database.

4.4.1.2 Discussion

Discussion objects comprised the `discussions` collection in the database.

1. `board_id` - the ID of the Board it is associated with.
2. `short_id` - the ID of the Discussion which should likely be unique among the Discussions associated with same the Board.
3. `chat` - the ordered list of Units in the chat section.
4. `summary` - the ordered list of Units in the summary section.
5. `focused` - the ordered list of Units in the focused section.
6. `id` - the unique `string` representation of the Discussion in the database. The concatenation of the `board_id` and `short_id`.

4.4.1.3 Unit

Unit objects comprised the `units` collection in the database.

Units were never truly removed from the database, only put in a hidden state, due to the complexities in Unit relationships that could arise if it was deleted, such as the subtle appearance of null references.

1. `board_id` - the ID of the Board it is associated with.
2. `short_id` - the ID of the Unit which should likely be unique among the Units associated with the same Board.
3. `pith` - the text within the Unit.
4. `position` - the x, y position of the Unit on the 2D map of the Board.
5. `in_chat` - whether the Unit was created through some Discussion chat (`true`) or the Board (`false`).

6. `author` - only applicable if the Unit came from the chat; the User that posted the Unit.
7. `created_at` - only applicable if the Unit came from the chat; the time the Unit was posted.
8. `hidden` - whether the Unit is hidden or not.
9. `id` - the unique `string` representation of the Discussion in the database. The concatenation of the `board_id` and `short_id`.

4.4.1.4 User

User objects comprised the `users` collection in the database.

Unlike in the alpha prototype, a user was associated with the board, the overall large group, instead of the small discussion. The name of the user was the same across all the small discussions they join.

1. `board_id` - the ID of the Board it is associated with.
2. `short_id` - the ID of the User which should likely be unique among the User associated with the same Board.
3. `discussion_id` - the Discussion the User is in if they are in a Discussion.
4. `nickname` - the displayed name of the user.
5. `id` - the unique `string` representation of the User in the database. The concatenation of the `board_id` and `short_id`.

4.4.1.5 Link

Link objects comprised the `links` collection in the database.

A Link encoded the relationship between one Unit and another and contained a description of the relationship. Links could be created on the board interface.

1. `board_id` - the ID of the Board it is associated with.
2. `short_id` - the ID of the Link which should likely be unique among the Links associated with the same Board.
3. `source` - the source Unit of the link relationship.
4. `target` - the target Unit of the link relationship.
5. `pith` - the description of the nature of the link.
6. `id` - the unique `string` representation of the Link in the database. The concatenation of the `board_id` and `short_id`.

4.4.1.6 Transclusion

Transclusion objects comprised the `transclusions` collection in the database.

A Transclusion encoded the relationship of one Unit transcluding another in its pith.

1. `board_id` - the ID of the Board it is associated with.
2. `short_id` - the ID of the Transclusion which should likely be unique among the Transclusions associated with the same Board.
3. `source` - The Unit with the transclusion.
4. `target` - The Unit being transcluded.
5. `id` - the unique `string` representation of the Transclusion in the database. The concatenation of the `board_id` and `short_id`.

4.4.2 Operations

The following specific operations were implemented to allow the given functionality specified in the design of the beta prototype. The IDs taken as input for all but the Board object were the `short_id` of the input object.

4.4.2.1 Board

The following operations could be performed in relation to a board.

A client needed a link to access the board. The link encoded the ID of the board.

The person must create a nickname for a board if they visited it for the first time and get back their ID within the board. The user ID would be stored in relation to the board ID in their longer storage. Thus, when they visited the board again, they would no longer need to create a user and could directly load the board directly.

1. `join_board(board_id)` - Tests that the given ID, most likely in the link the client received, is of an actual Board. If not, the frontend should render a 404 page that the link is invalid.
2. `create_user(board_id, nickname)` - creates the user ID given the nickname the user inputted for the board.
3. `load_board(board_id, user_id)` - returns the data the user needs to see the board: their nickname, the units on the board, and other information about the board.
4. `add_unit(board_id, text, position)` - adds a unit with the given text at the position on the 2D map of the board.
5. `remove_unit(board_id, unit_id)` - removes (actually hides) the unit from the board.
6. `edit_unit(board_id, unit_id, text)` - edits the unit with the new text for the pith.
7. `move_unit(board_id, unit_id, position)` - moves the unit to the new position.
8. `add_link(board_id, pith, source, target)` - adds a link between the source unit and the target unit with the pith description of the relationship.
9. `remove_link(board_id, link_id)` - removes the link.
10. `get_unit(board_id, unit_id)` - returns more detailed information about the unit, about which units it has links to and from and what discussions are on it.
11. `create_disc(board_id, unit_id)` - creates a discussion originating on the unit.
12. `search(board_id, query)` - searches through the units of the board with the given query. Can be useful for transclusions.

4.4.2.2 Discussion

The following operations could be performed in relation to a small discussion.

1. `join_disc(board_id, discussion_id, user_id)` - lets the user join the discussion. Returns information about the discussion: the units in the chat section, summary section, and focused section, as well as the list of participants. Emits to everyone in the discussion that the user has joined.
2. `leave_disc(board_id, discussion_id, user_id)` - lets the user leave the discussion. Emits to everyone in the discussion that the user has left.
3. `post(board_id, discussion_id, user_id, text)` - creates a post in the chat with the given text authored by the given user. Emits to everyone in the discussion about the new post.
4. `add_summary(board_id, discussion_id, unit_id)` - adds the chat unit to the summary section. Emits to everyone in the discussion about the update.
5. `remove_summary(board_id, discussion_id, unit_id)` - removes the chat unit from the summary section. Emits to everyone in the discussion about the update.
6. `add_focused(board_id, discussion_id, unit_id)` - adds the board unit to the focused section. Emits to everyone in the discussion about the update.
7. `remove_focused(board_id, discussion_id, unit_id)` - removes the board unit from the focused section. Emits to everyone in the discussion about the update.
8. `search(board_id, discussion_id, query)` - searches over the chat units in the discussion with the given query. Useful for transclusions.

4.4.3 Issues

4.4.3.1 Concurrency

Through our alpha implementation, we realized using locks to manage the editing and removing of units was too slow. We thus avoided using any concurrency mechanic on the board. However, this reintroduces the possibility that several users might try to edit or remove a unit or remove a link. Through further testing with real users, we may decide whether this is a serious problem.

4.4.3.2 Chat Loading

Loading all the posts became a slow operation as the number of posts grew. Therefore, we changed the system so that when a user first loaded an existing discussion, they only received the first 25 posts. Only when they scroll further back did they retrieve more posts from the server, 25 at a time.

4.4.3.3 Update Propagation

For the small discussion, it was relatively fast and efficient if every new update by one client was emitted to all other clients in the discussion. The small discussion was not meant to have many people at a time, and therefore the number of updates in a given period was expected to be relatively small.

However, the board was designed to be used by a larger group of people. Therefore, the number of updates to the board could be fairly high. Immediate pushes of each update would mean that each client may have to process a lot of updates to the board from other clients in a short period of time. The user themselves may become overwhelmed if the interface changed so much.

Thus, we used a pull mechanism to update the board. In this way, each client would periodically receive an update of the board, of the changes made since the last update.

The server used a chron job to mediate when to broadcast updates to the users of a board. Every 10 seconds, the server would loop through all the boards, and for each board, it would send out all the changes that occurred on the board to all the board's users since the last time it sent an update to them.

We therefore had to keep track of when changes were made since the last update. When users performed an operation on the board, the update-operation was recorded within a collection dedicated to recording updates. A timestamp was associated with the recorded operation. Thus, upon every update, the server could query for all the update-entries with a timestamp more recent than the last time it performed the update and send those.

4.5 Future

4.5.1 Test Suites

For the beta prototype, we performed testing in a mostly ad hoc fashion. Formal testing was done during the alpha phase, but it turned out to be more trouble than it was worth to maintain. However, as the codebase eventually stabilizes, it would be worth creating formal test suites to maintain assumptions about the behavior of operations.

4.5.2 Caching

Our system does not currently use caching between the app servers and the database. This could cause more severe latency and network usage issues as the number of users and user requests grow. It would be worth investigating a good cache solution, perhaps with Redis or Memcached, to speed up the retrieval of recently or commonly used data.

Chapter 5

Experiments

5.1 Research Questions and Hypotheses

In order to test the effectiveness of Pith, we conducted several studies to examine how a small group of people used the system when given a fixed topic of discussion for a short duration of time.

The main research questions we sought to answer through the studies were the following:

1. Is the system, its parts, and features understandable and usable to new users?
2. Does the system facilitate discussions that are productive?
3. Does the system help make diverse perspectives heard?

5.1.1 Usability

We had designed our system to be as simple as possible, with only the features we deemed most essential to achieving the goals of scale, productivity, and diversity highlighted in Chapter 3. Thus, we hoped that for the most part, the system would be understandable to users. For more unconventional features, such as the focusing section of the discussion, we anticipated that users would be less likely to use them, or use them in a way we would not have expected.

5.1.2 Productivity

Our system uses summarization to encourage productivity in two forms: first, in facilitating the creation of a takeaway product and second in promoting the quantity and quality of ideas through helping everyone keep track of what has been done and what needs to be done. Summarization appears mainly in two places: the summarization section for the discussion to organize chat posts and the board to organize the overall ideas. Thus, we anticipated that users would take advantage of these features to organize their ideas and identify where they could generate more ideas or to establish which ideas to reflect on and develop further.

5.1.3 Diversity

Board meta-cognitive features was our main way to address diversity of viewpoints, especially at a large scale. However, for these initial studies, we did not include as many features related to meta-cognition, such as beams or reflection periods. Part of this was because board meta-cognitive features are more relevant at a higher-level of discussion, which emerges with a large number of people or over a long period of time. Thus, given our limited setting, testing the effectiveness of these features would be difficult.

Instead, we focused on developing features that encouraged active listening or observing others over generating content, such as typing indicators, listening mechanics, and attention mechanics.

We did not plan to formally test the scalability of our system from either a computer systems or design point of view given the constraints of our budget and the difficulty of setting up such an experiment. We hope to conduct informal experiments to test this in the future.

5.2 Setup

5.2.1 Overview

Each group we tested with had a discussion on the Pith system centered around solving a task. After their discussion, each member filled out a survey asking about their experience of the discussion and the system itself.

In the full study, we had three variations, where each variation revealed the effect of the addition of a key part of the system. In the first variation, participants only had access to the chat component of the system. In the second variation, participants had access to the chat and summary components of the system. In the third variation, participants had access to all components of the system: chat, summary, and board. By comparing the second variation with the first, we could study the effect of adding the summary space on the resulting discussion. By comparing the third variation with the first and second, we could study how the presence of a board for organizing and networking ideas changed the nature of the discussion.

The task given to the participants was for them to collaboratively solve the Stranded problem: identify and rank ten items they would take with them as a group if all of them were stranded. We chose this task because it promoted creativity, encouraged teamwork, and was simple to explain.

5.2.2 Procedures

Participants were prescreened on Prolific for the following requirements:

1. was over the age of 18,

2. spoke English fluently, as our platform and study materials were in English,
3. had a high rating on Prolific (over 95% approval) to ensure a good quality of response,
4. was not in a previous study with us.

For each of the three variations in a full study, we created a corresponding session on Prolific to recruit for ten people. Once the session was made publicly available, Prolific participants could read the text of the study to determine if they were interested in joining. To encourage participants to join, we mentioned we would provide a $10 base compensation for their full participation with a potential $5 bonus.

A participant joined the session by clicking on the provided link, which would take them to the session website. First, they would see the consent form. Only if they signed the form could they continue. After the consent form, they were given instructions for how to conduct the rest of their session, then given a tutorial of how the system worked. If the current session was of the first variation, the tutorial would only describe the chat component. If it was of the second variation, it would describe both the chat and summary. If it was of the third variation, it would describe the chat, summary, and board. Once they were done with the tutorial, they would enter the waiting room and wait for the others before joining the discussion.

When it was time for the discussion to start, the participants were sent to the same small discussion on Pith. They were given 15 minutes to solve the Stranded problem. At the end of the discussion, they were redirected to the survey. The survey included questions in line with our three goals: how usable they thought the system was, how productive they thought the discussion was, and how much they felt they were heard during the course of the discussion. When the participant was done, they would be redirected to Prolific via a completion link.

To encourage participants to focus on the task during the course of the discussion, we told them ahead of time that they would be asked about their final ranking in response to the Stranded task during the survey and could gain bonus payment from answering the question more thoroughly. For each item that all members of the group reported was in the ranking with at least a sentence of justification, each member of the group would receive an additional 50 cents in bonus. There was a maximum of $5 bonus from ten such items.

To ensure the discussion would start in a timely manner relative to when each participant started their session from Prolific, we had it set to start sixteen minutes from when the first participant entered the session. Each participant who joined would see a timer that counted down how much time they had left until they had to be in the waiting room for the start of the discussion. We also closed access to the session from Prolific once about ten minutes had passed from the entry of the first participant so that even if a participant joined ten minutes in, they still had five minutes to reach the waiting room after signing the consent form, reading the instructions, and performing the tutorial. Instead of directly redirecting them to the discussion at the time of start, we gave them a ready button to click within one minute of its appearance. This was to ensure they were paying attention to the session after the waiting period and would engage in the discussion itself.

We required at least three participants to be present in the discussion. Thus, if fewer than three people were available to start, we would terminate the session instead of continuing on.

Refer to Appendix B for the recruitment text we used on Prolific, Appendix C for the consent form, Appendix D for the provided instructions, Appendix E for the tutorial given, and Appendix F for the full list of questions in the survey.

5.2.3 Testing System

We built the website used to host the sessions. There was a study interface participants would use for their session. In addition, there was an administrative interface for us to set the parameters of the session and examine which stage of the session each of the participants were at.

Figure 5.1 shows the administrative interface. At the top are controls to input parameters, such as which of the three variations the current session is. Buttons on the right allow the admin to control the exact moment the ready button is presented to the participants and when to start and end the discussion. The columns at the bottom indicate the stages the participants could be in during the course of their session:

1. cancel – when the participant's session was canceled because they did not reach the waiting room in time, they did not click on the ready button in time, or there were not enough people to begin the discussion,
2. consent – when the participant could see the consent form,
3. instructions – when the participant could see the instructions,
4. tutorial – when the participant could see the tutorial,
5. waiting – when the participant was in the waiting room,
6. ready out – after the administrator clicked "Ready", when the participant should see the ready button,
7. ready in – when the participant had clicked on the ready button to indicate they were ready for the discussion,
8. discussion – when the participant was in the discussion,
9. survey: task – when the participant was taking the first portion of the survey which asked about the discussion,
10. survey: pith – when the participant was taking the second portion of the survey which asked about the Pith system,
11. complete – when the participant was done with their session and could return to Prolific.

Figure 5.2 shows what the key stages look like to the participant. The participant saw three timers during the course of their session:

1. While they were signing the consent form, reading the instructions, and following the tutorial, the first timer counted down when they should be in the waiting room. The timer started ticking from when the first participant entered their session and went on for 15 minutes. All participants saw the same amount of remaining time.

2. Once they received the ready button, the second timer counted down how long they had to press it. They were given a minute.
3. Once they were in the discussion, the third timer indicated when the discussion would be over. It ticked for 15 minutes.

The timer turned red once there was a minute left to warn participants. The administrator could also see the same timer.

Participants also had access to the instructions and tutorial during their time in the waiting room and the discussion to refresh their memory whenever needed. Figure 5.3 shows what the pop-ups for each appear as. In addition, Figure 5.3 shows the screen presented should a participant complete their session and the screen presented should their session be terminated. The session could be terminated if the participant did not reach the waiting room in time, if the participant did not click on the ready button in time, or if there were not enough participants in the session to continue to the discussion stage.

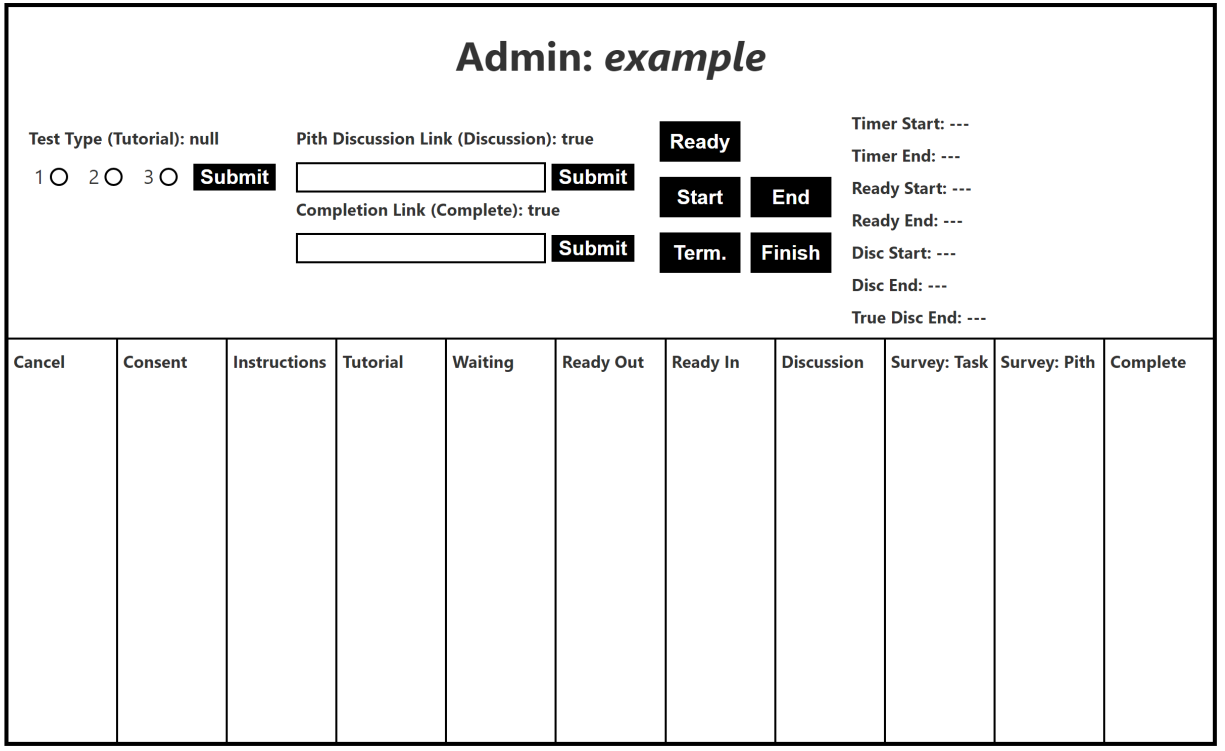


FIGURE 5.1: Screenshot of administrative interface.

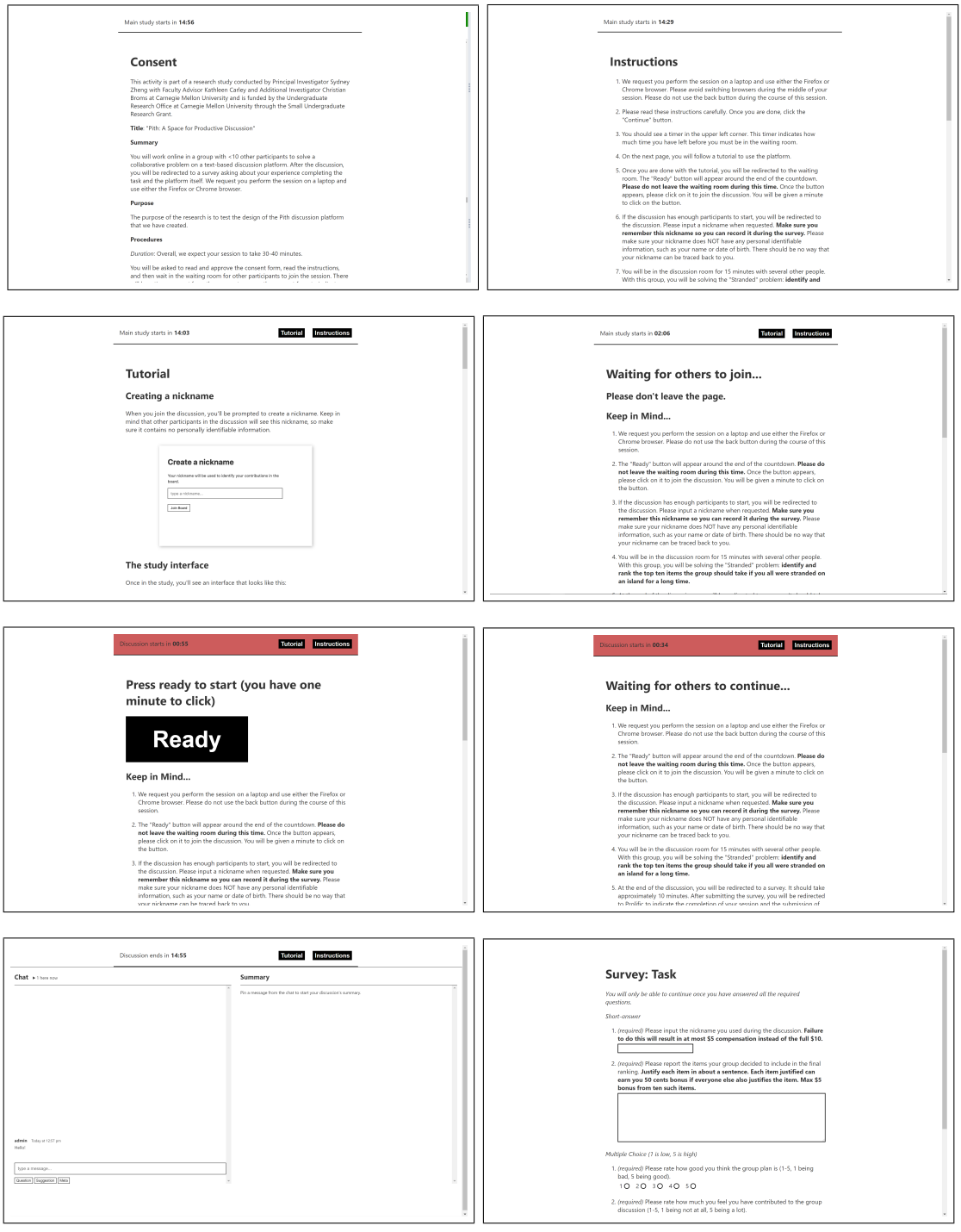


FIGURE 5.2: Screenshots of key stages in a participant's session. From left to right, top to bottom, the stages are consent, instructions, tutorial, waiting room, ready out, ready in, discussion (shown with the second variation with chat and summary), and survey.

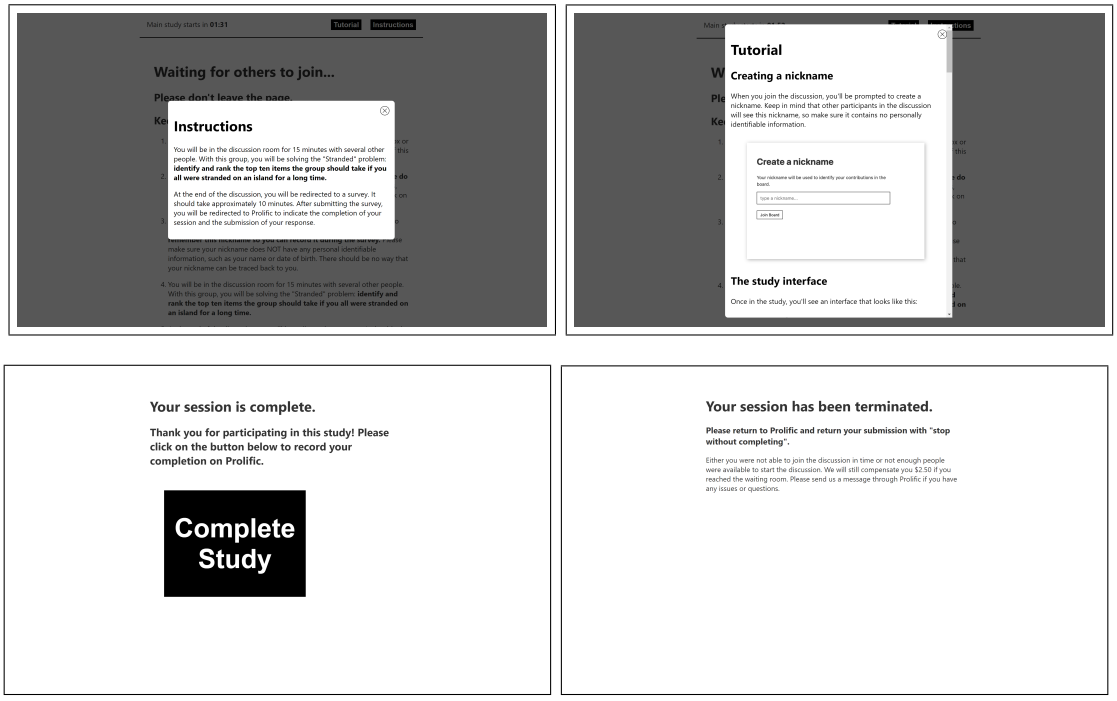


FIGURE 5.3: Screenshots of auxiliary shots in a participant's session. From left to right, top to bottom are the instruction pop-up, the tutorial pop-up, the complete screen, and the terminated screen.

5.3 Observations

5.3.1 Pilot

We first conducted a pilot study that followed the second variation with chat and summary. Our goal for the pilot was to identify any issues with the study setup before we conducted the full study.

In addition to the standard requirements, participants had to have participated in 100-10,000 Prolific studies. This was to ensure that participants had a good understanding of how studies worked in general and could provide feedback on the study itself. Indeed, the participants made several suggestions we found to be helpful. For example, one of them suggested the website should indicate when there was less than a minute left on the timer. We incorporated this feedback by turning the timer Indian-red once a minute was left, as can be seen in two of the sub-figures in Figure 5.2.

We ended up having eight participants complete their session. The pilot study gave us a good, initial sense of how participants could use our platform.

5.3.1.1 Productivity

The participants used the summary section to organize their ranking for the Stranded problem, as can be seen in Figure 5.4. In the survey, three of the eight participants noted how useful the summary was in organizing their ideas. One person mentioned that the summary was "a good way not to get lost in what you are talking about." This was encouraging, as one of our hypotheses was that the summarization feature allowed for increased productivity. Indeed, several of the participants copied the resulting ranking on the summary onto an external aid, such as on a digital notepad, in order to exactly answer the survey question about what their ranking was.

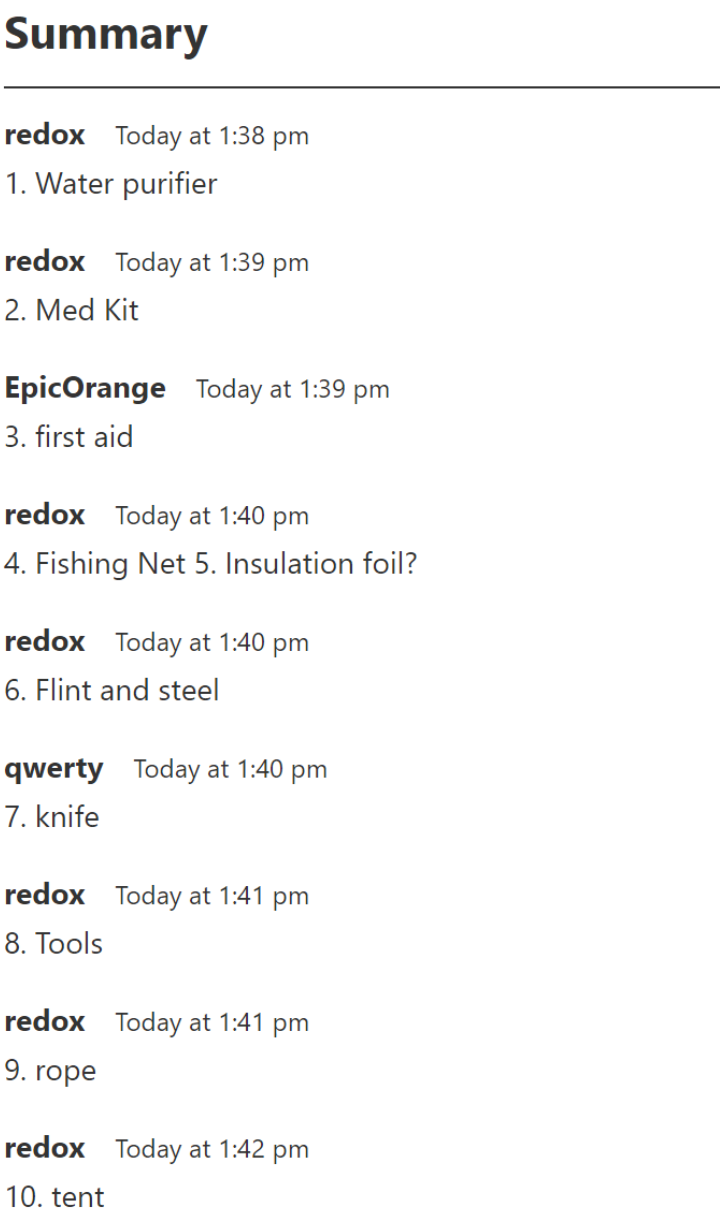


FIGURE 5.4: The summary with final Stranded ranking created by the pilot group.

5.3.1.2 Usability

Many participants commented on how simple the system was. Some of the comments were in a more positive light: "visually clear and uncluttered", "easy to understand", "clear and universal". Some were more negative, indicating the platform was "too simple" or "too plain". However, we generally took this to be a good sign that the interface was easy to use, despite the short window of time given for them to use it. Indeed, the majority of participants, 6 out of the 8, rated the system as being 8 or more on a scale of 1 being hard to use and 10 being easy to use. Refer to Figure 5.5 for the full breakdown.

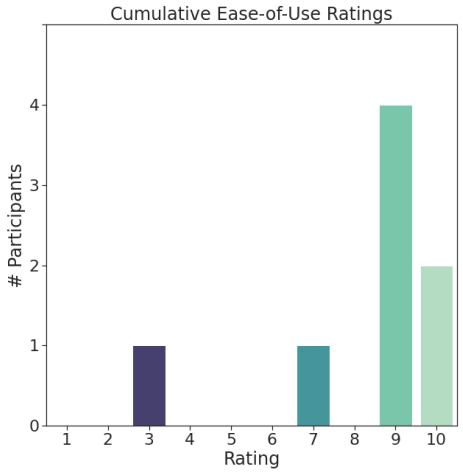


FIGURE 5.5: Ratings of how easy the people thought the system was to use (1 being hard, 10 being easy).

5.3.1.3 Diversity

When asked how much they felt they had made a contribution with 1 being not at all and 5 being a lot, 2 out of the 8 rated less than a 3. When asked how much they felt heard with 1 being not at all and 5 being a lot, 2 out of the 8 rated less than a 3. These numbers suggested that the system did not provide means for people to feel as if they were being heard or making a contribution. Refer to Figure 5.5 for the full breakdown.

We observed there were cases where participants would make a suggestion in a post and yet others would not respond to the post in any way. Additionally, several of the participants indicated on the survey that they felt they did not contribute as much to the discussion or did not feel heard. This point inspired us to address diversity more. We eventually came up with the idea of having post-flairs. Users could mark their posts as being a "Question" if it was a question, a "Suggestion" such as a suggestion for the summary, or "Meta" for high-level commentary, such as to flair a post encouraging the group to finish adding contributions and start ranking them. Our hope was that the flair-mechanic could draw people's attention to a post and encourage them to acknowledge or respond to it in some way. Namely, a person could

answer someone's question, the group could deliberate on a suggestion, and the group could change their discussion behavior in response to a meta.

We also had feedback about improving the interface itself. One person recommended that we include a feature to see who added or removed a post from the summary, which we implemented. Another recommended that posts on the summary could be moved around, which at the time, they could not. While this was a reasonable idea, we decided not to implement it at the time. We wanted to determine if this need was still present when participants used the third variation of the system; the third variation included the board, which allowed for visual positioning of units. There were some additional minor suggestions, which we included if possible.

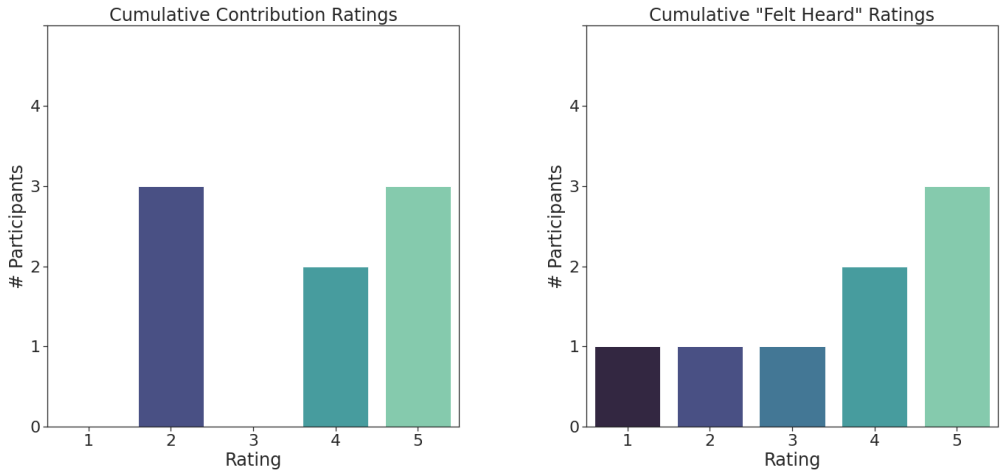


FIGURE 5.6: On the left, ratings of how much people felt they contributed to the discussion (1 being not at all, 5 being a lot) and on the right, ratings of how much they felt heard (1 being not at all, 5 being a lot).

5.3.2 Full Trial 1

For the first full trial, we decided to focus on conducting variation 1 and variation 2. The variation 1 session had 7 participants, who each gave a response. We ended up performing the variation 2 session twice because the first session encountered technical difficulties. However, we report on both sessions' results. The first variation 2 session had 5 participants with 4 responses, and the second had 8 participants with 7 responses. In the former case, a participant was not able to send us their response. In the latter case, a participant had an incomplete response and did not attempt to contribute to the discussion.

We did not conduct a variation 3 session because we believed it would be best to first have the small discussion system address the points of usability, productivity, and diversity before adding the board.

Our main two questions during this trial were the following:

1. Does the addition of the summary improve productivity?

2. Does the post-flair mechanic (described above) help address diversity by encouraging active listening of contributions?

5.3.2.1 Productivity

The experiments conducted in this trial suggested that the summary did improve productivity. A user in the variation 1 session mentioned that at least one person had to perform the role of compiling the suggested items. A few participants ended up performing this duty, likely with an external aid, such as a digital text processor. One participant even shared the bright idea that everyone should copy-paste the compiled ranking into a text editor so that all of them win the full bonus through answering the corresponding survey question completely. Indeed, everyone appeared to do this, and they won the full bonus.

In addition, two of the members noted that the structure of the system was not conducive to this sort of task. One person suggested that "a list that can be changed as needed is a good starting point", and the other observed that though the list was generated quickly, figuring out the priorities, or ranking of items was more difficult given the current format.

In contrast, users in the variation 2 groups had access to the summary section. The first variation 2 group only pinned the final ranking, generated purely through the chat. The second variation 2 group pinned generated ideas or intermediate forms of the ranking before pinning the final ranking at the end. No one in either sections complained of a lack of organization.

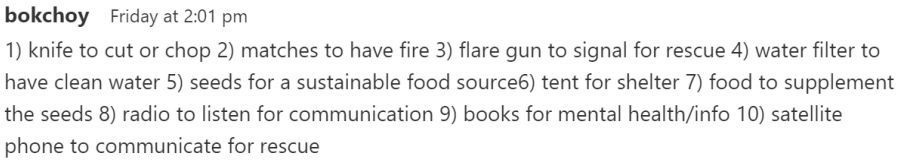


FIGURE 5.7: The final Stranded ranking compiled by the full study 1, variation 1 group.

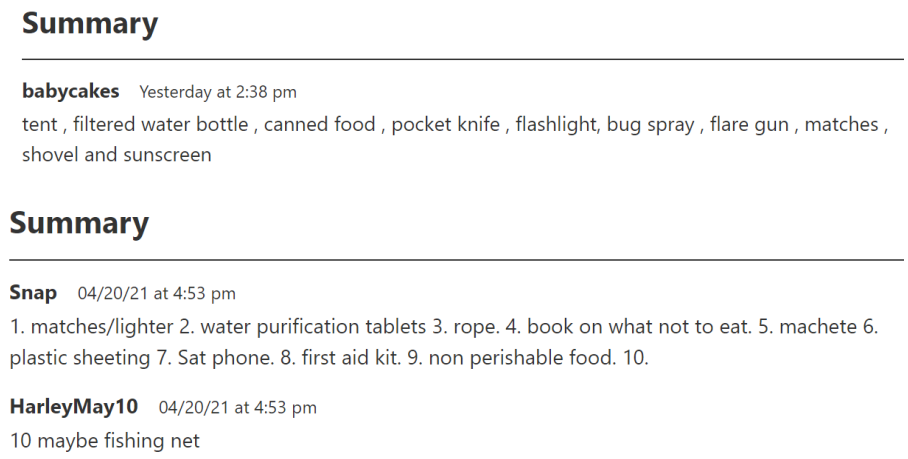


FIGURE 5.8: The summaries with final Stranded ranking created by the two full study 1, variation 2 groups.

5.3.2.2 Diversity

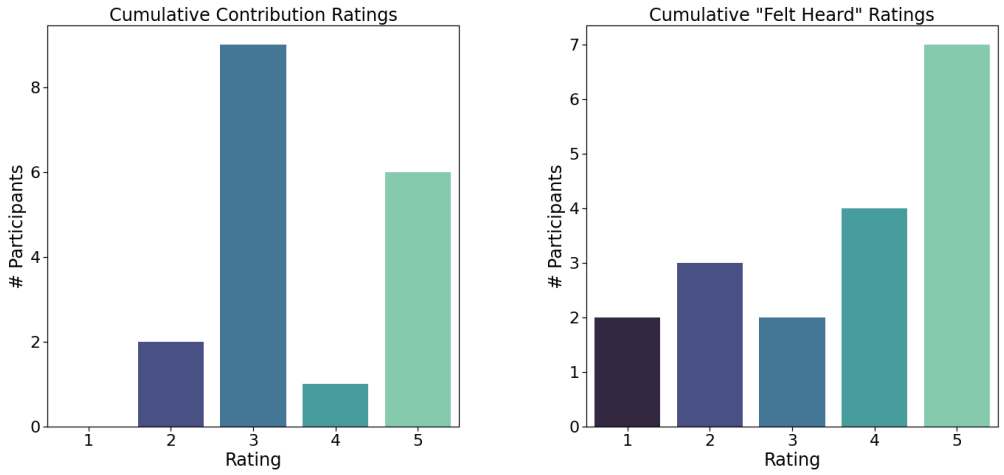


FIGURE 5.9: On the left, ratings across all three sessions of how much people felt they contributed to the discussion (1 being not at all, 5 being a lot), and on the right, ratings of how much they felt heard (1 being not at all, 5 being a lot).

We also examined each of the group's use of the unit flairs, either in the presence of a summary section for variation 2 groups or not, for the variation 1 group.

In the first variation 1 session, the "Suggestion" flair was used 3 times. Twice to propose a new item, which did not get an immediate response. Once to propose users to copy-paste the final ranking to get the full bonus, which others responded to affirmatively. The "Question" flair was used twice. Once to ask a clarification question and did not get an immediate response. The other instance was to elicit feedback from a specific user, which did get responded to immediately.

In the first variation 2 session, the "Suggestion" flair was only used once to propose an item. However, others did not respond to the proposal. The "Question" flair was used once to ask a clarification question, which was immediately responded to.

In the second variation 2 session, the "Suggestion" flair was used once, probably as a test of the feature as it was associated with a message that said "hi". The "Question" flair was used once to ask about an item with a certain property, but did not get immediate response.

Thus, whether the summary section was present or not, the post flair feature was used far less than we hoped. In particular, we intended the "Suggestion" flair to be used every time a user suggested a new item to the ranking, a reorder of the ranking, or an addition or removal to the summary board. The "Meta" feature was not used at all, but this is understandable given its less concrete nature. In addition, it did not necessarily encourage a response anymore than a normal post. It seemed whether it elicited a response was more dependent on the actual content, such as the case where it was addressed to a specific person or promoted a popular idea.

It is also apparent that overall, users still did not feel as if they were being heard. On the question asking them how much they felt they contributed to the discussion, where 1 is not at all and 5 is a lot, 11 out of 18 people across all three groups rated less than a 4. On the question asking them how much they felt heard, where 1 is not at all and 5 is a lot, 7 out of 18 people across all three groups rated less than a 4. Refer to Figure 5.9 to see the full breakdown of ratings for both questions.

One user mentioned that they wished there had been some mechanism for them to catch the others' attention and have them notice the user's contributions.

Our proposal to improve the system given the feedback was to increase the stakes when using flairs. There was no incentive for users to take notice of a post, even if it is flaired. We therefore proposed that when a user flaired their post with any option (i.e. "Question", "Suggestion", "Meta") it should automatically be pinned to the summary and generate a notice unit, broadcasting the action. These means should further draw people's attention to the post's contribution. Users should be encouraged to address the post, lest it clutter their summary space. Once they have addressed the post, they could finally remove it from the summary section. We hoped users, in seeing the behavior of flairing, would feel incentivized to use it more often.

5.3.2.3 Usability

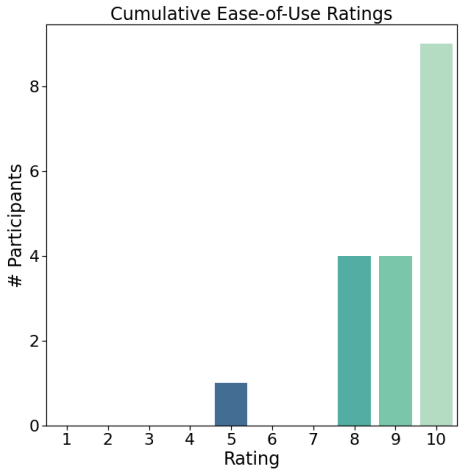


FIGURE 5.10: Ratings across all three sessions of how easy the people thought the system was to use (1 being hard, 10 being easy).

As before, users mentioned the simplicity and ease of the system, at least for the most part: 1 out of 7 in the variation 1 group, 2 out of 4 in the first variation 2 group, 3 out of 7 in the second variation 2 group). When the participants rated the ease-of-use of the system overall, only one person in all three sessions of this trial rated below an 8, where 1 is hard and 10 is easy. This participant was in the variation 1 group and gave a rating of 5, and was one of the two participants who complained the system did not allow for sufficient organization of ideas. Refer to Figure 5.10 to see the full breakdown of ratings.

There were a few issues people noted about the system that prevented perfect ease-of-use. Similar to before, one person believed that clicking on a post should copy its link directly to the text-space for a new post rather than only adding it to the clipboard for them to manually paste in the text-space. Another person had difficulty finding the list of participants feature, which we implemented after the pilot study. One person was displeased that others had the ability to unpin their pins, a reasonable complaint.

We also noticed cases where people would respond to certain people. This was despite the presence of the reply-to feature. It is possible that there are cases people would desire to use a mention feature (i.e. @user1) over a reply to a specific post. However, perhaps in its current form, the reply-to is too difficult to use. Once we are able to make the reply-to feature easier to use by having it automatically paste a link to the text-box to create a new post, it would worth examining if the desire to mention certain people still exists and is a needed feature in its own right.

5.3.3 Full Trial 2

We conducted one variation 1 session and one variation 2 session. The variation 1 session had 8 participants with 8 responses. The variation 2 session had 7 participants, with 6 responses; we threw out one set of responses because it appeared the participants was having technical difficulties with the system that prevented them from participating in the discussion. Our primary purpose for this set of studies was to first, determine if our modification on flairs improved the usage of flairs and the diversity ratings, and second, whether the observation that a summary helped organize conversation persisted.

5.3.3.1 Diversity

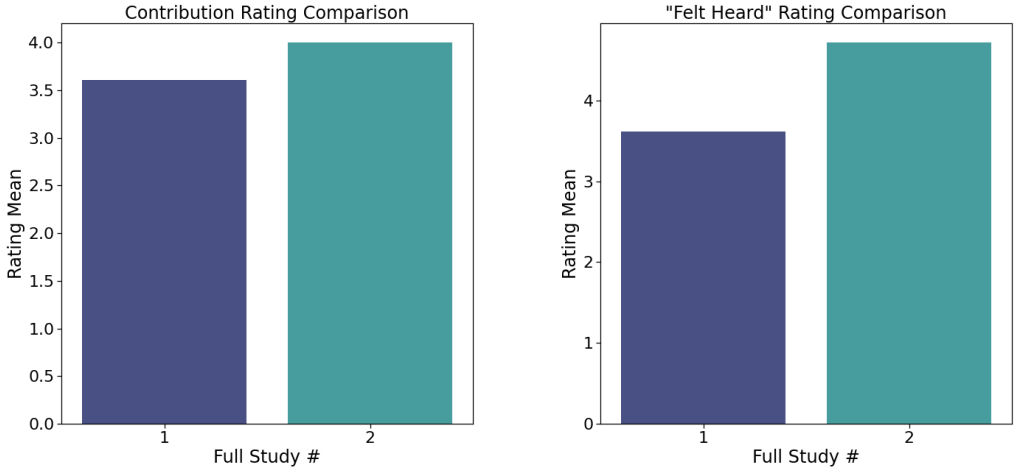


FIGURE 5.11: On the left, a comparison of the means of the contribution rating from the first and second full study participants. On the right, a comparison of the means of the "felt heard" rating from the first and second full study participants.

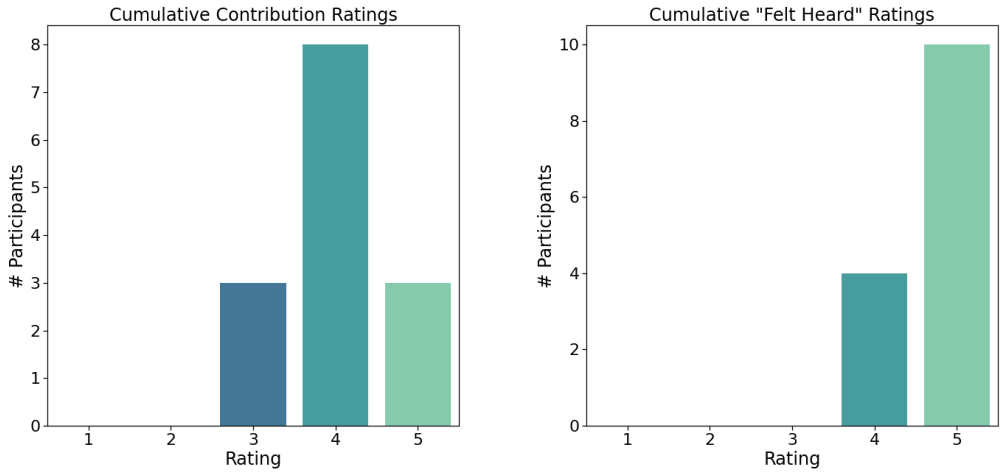


FIGURE 5.12: On the left, ratings across the two sessions of how much people felt they contributed to the discussion (1 being not at all, 5 being a lot), and on the right, ratings of how much they felt heard (1 being not at all, 5 being a lot).

Our most important general observation of both sessions was that they used the flairs more–most participants in either group used the suggestion flair at least once to share an item. In addition, 2 out of 6 people in the second variation group and 3 out of 8 people in the first variation group mentioned that they liked the flair feature. In the case of the variation 1 group, their setup did not change from the variation 1 group in the previous full study. However, in the case of the variation 2 group, they may have been further incentivized to use the flair feature because it automatically pinned the post to the board.

The distribution of participant ratings for "feeling heard" and for feeling that they made a contribution subjectively increased, as can be in seen in Figure 5.11, comparing the means of the two questions as gathered from the first full study and the current full study. The full distribution for each question is shown in Figure 5.12. This indicates promise for the flair feature as helping people feel they are heard or for promoting active listening. However, it is certainly not a conclusive result–it is possible the higher ratings were due to other circumstances related more to this particular group of participants than their usage of the flair feature.

5.3.3.2 Productivity

The variation 1 group used flairs to organize their conversation. Indeed, their group was the only one that used the "Meta" flair, which they used to highlight the current ranking, as presented in Figure 5.13. One participant mentioned that someone had to compile the current ranking, which was mentioned in the previous variation 1 study group. However, generally speaking, this group had little complaints about achieving the task in an organized manner.

In the variation 2 session, we observed a continued trend of people greatly preferring the summary, with all 6 people making some positive reference to it in the survey. The final ranking was still saved in the summary section as before. In Figure 5.14, it is the fifth post from the

bottom. However, perhaps because the flairs automatically added posts to the summary, the summary grew more cluttered with posts. The participants also never unpinned a post. This appeared to cause confusion for at least one participant, who noted that it was unclear when to use the chat section and when to use the summary section. Likely because of this, she assigned the single ease-of-use rating of 6. Perhaps the considered use of unpinning could help maintain the succinctness of the summary section.

It is possible that people did not use the unpinning feature because the flair on many of the posts implied to them it should stay. However, most posts that are pinned through a flair are only important temporarily, not indefinitely. Once the question is asked or the suggestion considered, the post may no longer be relevant. One idea we had to prevent clutter was to automatically unpin posts on the summary originally pinned because of a flair. In this case, it should be easy for people to repin the post if they believed it was still worth preserving.

However, people should feel comfortable unpinning posts themselves. One of our ideas to encourage considered unpinning use was to have a warning feature. Every time someone unpins a post from the summary, a banner can appear at the top, asking if the user remembered to consult with the author of the post or the pinner before they unpinned the post. Hopefully through such a reminder, a culture would be established among users that the unpin feature should be used, but with proper consideration.

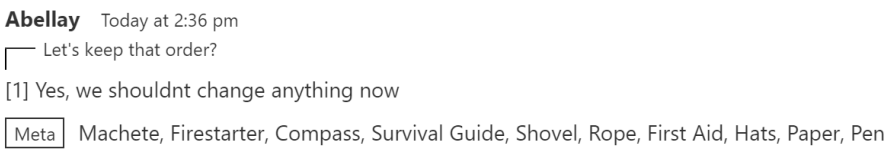


FIGURE 5.13: The final Stranded ranking compiled by the full study 2, variation 1 group, which is marked by a "Meta" flair.

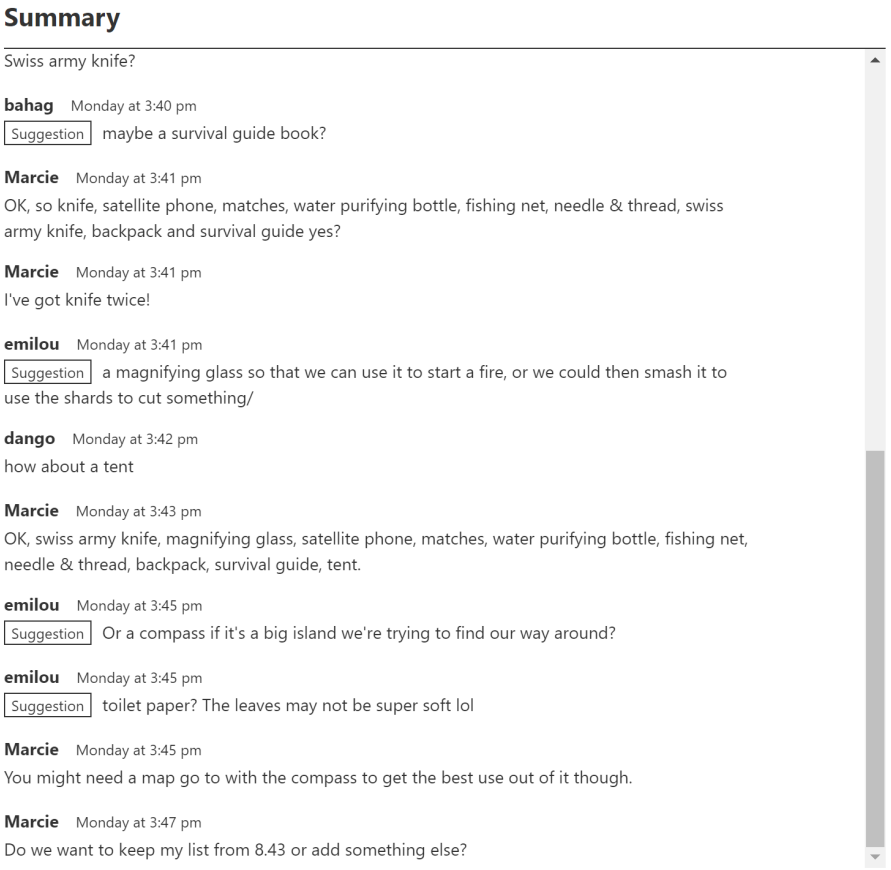


FIGURE 5.14: The summary section in its final form from the study 2, variation 2 group.

5.3.4 Usability

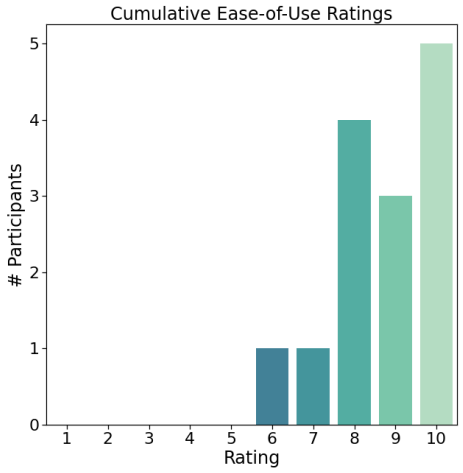


FIGURE 5.15: Ratings across the two sessions of how easy the people thought the system was to use (1 being hard, 10 being easy).

Generally speaking, the ratings for ease-of-use were still fairly high, with 2 out of 14 participants rating under an 8. The most common complaint, mentioned by at least four participants, was that the reply-to was unintuitive to use because it required copying and pasting the post ID into the text-box rather than pasting the post ID automatically. We received a thoughtful suggestion from one participant, who noted that to make the typing indicator even more useful in letting others know who was typing, a color could be assigned to each participant and the color could be displayed as part of the typing indicator. We plan to integrate this idea in a future version of our prototype.

Chapter 6

Conclusion

6.1 Limitations

Our goal was to develop a minimum viable product of our system. While we finished most of the implementation, there were parts of the system that were considerably rough. In particular, because we built the 2D map interface of the board from scratch instead of using pre-existing libraries in a few weeks, the graphics did not move around naturally. We avoided using large pre-existing libraries because the server the system ran on had limited memory. In addition, we believed that using light-weight libraries was generally a good practice. Otherwise, there were seemingly minutiae bugs that noticeably affected the user experience that we did not discover until later and thus did not have time to address.

While it was necessary for us to have sufficient time to develop our minimum viable product in the first place, to have a usable system required periodic user feedback. Because of our timeline, we had only about one month to conduct experiments. Due to the complexity of our study setup, we had to build our own study website. Participants experienced bugs, which took a while to fix. This meant that a session would have to be repeated several times. For example, we ended up conducting the full study 1, variation 2 study type six times in order to reach a working session. After the pilot and each full study, we analyzed the data to determine how to improve our system.

We decided to dedicate our testing to the the small discussion part of the system, and so only worked with variation 1 and 2 study types. We desired to address the goals of usability, productivity, and diversity before adding an additional layer of complexity, as introduced by the board in the variation 3 study type. Because of this, we did not spend as much time finishing the implementation of reflection periods, a key meta-cognitive board feature.

We were encouraged to design the study so a participant's session would be less than an hour, ideally around half an hour. Thus, the participants only had 15 minutes for a discussion, as some time was needed for them to fill a survey and gather in the waiting room. Because of the limited time people had to interact with the system, we did not reveal some of the features it

had, such as the search functionality. In addition, the task chosen had to be relatively straightforward. Thus, the product had to be fairly simple, and indeed for the Stranded problem, the discussion product was a fixed-size ranking. However, for real use-cases, the generated products could be more complicated, in ways that we cannot currently anticipate but that we eventually want to study.

6.2 Future Work

We hope to continue working on the Pith system beyond this thesis. Before anything else, we would like to continue our studies by performing the variation 3 study type with several groups to understand the basic usage of our current system.

We would improve the system design by, at the very least, implementing a caching layer between the app and database layer. In addition, we should automatically delete intermediate data from the database that could be deleted.

The next main step would then be to deploy our system to the world. We would have a feature to allow real users to provide feedback about the system. In addition, we may conduct analyses on the discussions of organic users with their approval to understand the diverse use cases of the system. To further the diverse use of our system, we should aim to provide wider language support. Currently, the system is built for English-users. Finally, over time, we hope to expand our vision of the Pith system by integrating the ideas mentioned in the two "Other Ideas" subsections of the "Design" Chapter.

6.3 Conclusion

We conceived, designed, and implemented the key design of our experimental Pith system to facilitate large-scale, productive discussion.

To allow participants of the discussion experience natural conversation, we proposed and integrated the idea of breaking the large group up into smaller discussion groups.

To encourage productivity within the discussion, our system featured two levels of summarization: a summary section within the small discussion interface to help a small group keep track of their discussion, and a board to allow for intercommunication among the small discussion groups and to help the overall group visualize the state of their overall discussion.

To encourage diversity on our platform, we implemented group meta-cognitive mechanisms, beams and reflection periods in particular. Beams allowed users to communicate high-level suggestions about the direction of conversation or the visual organization of the board. Reflection periods allowed the group to spend some time away from generating ideas to organizing the ideas and planning the future course of their discussion. In addition, we proposed and tested active listening mechanisms on the small discussion side. Additionally, we avoided the commonly-used

popularity metrics to encourage people to judge ideas critically on their own rather than relying on group-think.

Our experiments were focused on studying how usable our system was, how much it encouraged productive discussion, and how much it encouraged diversity, which in a small group discussion was achieved through participants actively listening to each other. Our first main result was that our system was easy to use, at least with a short tutorial. The second result was that the summary section appeared to help the group organize their thoughts and form a working product members can take away. The third was that our flairs mechanic was promising in encouraging active listening among participants.

Appendix A

Packages and Tools Used

A.1 Backend Tools

1. Docker [67]
2. Python 3.8 [73]
3. HAProxy [64]
4. Ubuntu 18.04 [74]

A.2 Frontend Tools

1. Svelte [70]
2. Sapper [71]
3. React [68]
4. Redux [69]
5. node.js [75]

A.3 Docker Images

image	Docker Hub
node.js [75]	Docker link
haproxy [64]	Docker link
Ubuntu 18.04 [74]	Docker link

A.4 Key Javascript Packages

The following were used for the frontend.

package	author	source
day.js	"iamkun"	Github link
panzoom	Andrei Kashcha	Github link

A.5 Python Packages

The following were used for the backend. Missing values in the table below indicate they were not present.

package	author	PyPI	source
aiodns 2.0.0	Saúl Ibarra Corretgé	PyPI link	Github link
aiohttp 3.7.4.post0	Nikolay Kim	PyPI link	Github link
aioredis 1.3.1	Alexey Popravka	PyPI link	Github link
aniso8601 7.0.0	Brandon Nielson	PyPI link	Bitbucket link
arq 0.19	Samuel Colvin	PyPI link	Github link
async-timeout 3.0.1	Andrew Svetlov	PyPI link	Github link
attrs 20.1.0	Hynek Schlawack	PyPI link	Github link
cchardet 2.1.6	"PyYoshi"	PyPI link	Github link
certifi 2020.6.20	Kenneth Reitz	PyPI link	Github link
cffi 1.14.2	Armin Rigo, Maciej Fijalkowski	PyPI link	Heptapod link
chardet 3.0.4	Mark Pilgrim	PyPI link	Github link
dnspython 1.16.0	Bob Halley	PyPI link	Github link
eventlet 0.26.1	Linden Lab	PyPI link	Github link
greenlet 0.4.16	Ralf Schmitt	PyPI link	Github link
gunicorn 20.0.4	Benoit Chesneau	PyPI link	Github link
h11 0.9.0	Nathaniel J. Smith	PyPI link	Github link
hiredis 1.1.0	Jan-Erik Rediger, Pieter Noordhuis	PyPI link	Github link
httptools 0.1.1	Yury Selivanov	PyPI link	Github link
idna 2.10	Kim Davies	PyPI link	Github link
ipaddress 1.0.23	Philipp Hagemeister	PyPI link	Github link
iso8601 0.1.12	Michael Twomey	PyPI link	Github link
joblib 0.15.1	Gael Varoquaux	PyPI link	Github link
jsonschema 3.2.0 [72]	Julian Berman	PyPI link	Github link
mongoengine 0.20.0	Harry Marr	PyPI link	Github link
monotonic 1.5	Ori Livneh	PyPI link	Github link
mypy 0.782	Jukka Lehtosalo	PyPI link	Github link
mypy-extensions 0.4.3	Jukka Lehtosalo	PyPI link	Github link

promise 2.3	Syrus Akbary	PyPI link	Github link
pycares 3.1.1	Saúl Ibarra Corretgé	PyPI link	Github link
pycparser 2.20	Eli Bendersky	PyPI link	Github link
pydantic 1.6.1	Samuel Colvin	PyPI link	Github link
pyflakes 2.3.1		PyPI link	Github link
pymongo 3.10.1 [66]	Bernie Hackett	PyPI link	Github link
pyrsistent 0.17.0	Tobias Gustafsson	PyPI link	Github link
python-dateutil 2.8.1	Gustavo Niemeyer	PyPI link	Github link
python-dotenv 0.14.0	Saurabh Kumar	PyPI link	Github link
python-engineio 4.0.0	Miguel Grinberg	PyPI link	Github link
python-socketio 5.0.4 [63]	Miguel Grinberg	PyPI link	Github link
pytz 2020.1	Stuart Bishop	PyPI link	Github link
redis 3.5.3 [65]	Andy McCurdy	PyPI link	Github link
regex 2020.6.8	Matthew Barnett	PyPI link	Bitbucket link
requests 2.24.0	Kenneth Reitz	PyPI link	Github link
Rx 1.6.1	Dag Brattli	PyPI link	Github link
sentinels 1.0.0	Rotem Yaari	PyPI link	
singledispatch 3.4.0.3	Jason R. Coombs	PyPI link	Github link
six 1.15.0	Benjamin Peterson	PyPI link	Github link
typed-ast 1.4.1	David Fisher	PyPI link	Github link
typing-extensions 3.7.4.3	Guido van Rossum, Jukka Lehtosalo, Lukasz Langa, Michael Lee	PyPI link	Github link
urllib3 1.25.10	Andrey Petrov	PyPI link	Github link
uvicorn 0.11.7	Tom Christie	PyPI link	Github link
uvloop 0.14.0	Yury Selivanov	PyPI link	Github link
watchgod 0.6	Samuel Colvin	PyPI link	Github link
websocket-client 0.57.0		PyPI link	Github link
websockets 8.1	“liris”	PyPI link	Github link
yarl 1.5.1	Andrew Svetlov	PyPI link	Github link

Appendix B

Recruitment

The following text, approved by the IRB, was posted on Prolific to recruit potential volunteers of our study.

Title: "Pith: A Space for Productive Discussion"

This is a research study conducted at Carnegie Mellon University. Participation is voluntary and appreciated.

(only shown for pilot study) **Note**: This is a pilot study, so there may be technical difficulties.

Requirements: age 18+, fluency in English, Prolific approval rating >95%, have not been in a previous study with us.

(only shown for pilot study) We also require that you have done between 100-10000 Prolific studies previously.

Purpose: The purpose of the research is to test the design of a system we created, the Pith discussion platform.

Summary of Procedures: You will work online in a group with <10 other participants to solve a collaborative problem on a text-based discussion platform. After the discussion, you will be redirected to a survey asking about your experience completing the task and the platform itself. Overall, we expect your session to take 30-40 minutes. We request you perform the session on a laptop and use either the Firefox or Chrome browser.

Payment: You will be compensated $10 if you complete all parts of the study (i.e., sign consent form, read instructions and do tutorial, perform discussion, complete survey). There is a possible bonus of up to $5.

Please note that if either of the following two scenarios occur, you will be paid $5.00 instead of $10, and there will be no possibility of a bonus:

1. If you complete the discussion but do not complete the required questions of the survey.

2. If you complete the discussion but do not report the valid username (nickname) you used in the discussion on the survey.

Please note that if either of the following two scenarios occur, you will be paid $2.50 instead of $10, and there will be no possibility of a bonus:

1. If you sign the consent form, but there are not enough people to participate in the discussion alongside you.
2. If you sign the consent form but do not reach the discussion in time and therefore cannot complete the discussion and after-discussion survey.

If you do not sign the consent form, there is no chance of any compensation, partial or full.

Contact: Email Principal Investigator Sydney Zheng at slzheng@andrew.cmu.edu.

Appendix C

Consent Form

The following form, approved by the IRB, was presented to potential participants before they began in our study.

This activity is part of a research study conducted by Principal Investigator Sydney Zheng with Faculty Advisor Kathleen Carley and Additional Investigator Christian Broms at Carnegie Mellon University and is funded by the Undergraduate Research Office at Carnegie Mellon University through the Small Undergraduate Research Grant.

Title: "Pith: A Space for Productive Discussion"

Summary

You will work online in a group with <10 other participants to solve a collaborative problem on a text-based discussion platform. After the discussion, you will be redirected to a survey asking about your experience completing the task and the platform itself. We request you perform the session on a laptop and use either the Firefox or Chrome browser.

Purpose

The purpose of the research is to test the design of the Pith discussion platform that we have created.

Procedures

Duration: Overall, we expect your session to take 30-40 minutes.

You will be asked to read and approve the consent form, read the instructions, and then wait in the waiting room for other participants to join the session. There will be a timer present from the moment you see the consent form to indicate how much time you have left until you should be in the waiting room. You should have between 5-15 minutes.

If you are in the waiting room, you will receive a button that you can click on to join the discussion around the time the timer finishes. You must click the button within a minute after

seeing it. **If you do not enter the waiting room in time and click on the button, you will not be able to progress further in the study.**

You will have exactly 15 minutes to work with the others in the discussion on the Pith platform to solve the Stranded problem. The goal is to identify and rank 10 items you would take with you as a group if all of you are stranded.

Once the 15 minutes are up, you will be redirected to a survey, which you should be able to complete in about 10 minutes. Once you are done, you will be redirected to the Prolific site and your session is complete.

NOTE: In the case that there are fewer than three people that are ready to start the discussion, the discussion will also not continue. The study will end early, and if you have clicked the button in the waiting room, you can be redirected to Prolific.

Participant Requirements Participation in this study is limited to individuals who are age 18 and older, can speak English fluently, have a Prolific approval rating >95%, and have not been in a previous study with us.

(only shown for pilot study) We also require that you have done between 100-10000 Prolific studies previously.

Risks

The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or while interacting with others in online activities.

Benefits

Participants may enjoy using and contributing feedback about a novel discussion platform aimed at productive large-scale discussion.

Compensation & Costs

Base compensation

If you sign the consent form, participate in the discussion, and complete the survey, you will receive the full payment of $10.

Bonus

After the discussion and in the survey, you will be asked to report the items your group decided to include and say why you've included each item. For each item that *all* members of the group reported was in the ranking with at least a sentence of justification, each member of the group will receive an additional 50 cents in bonus. Maximum $5 bonus from ten such items.

If you are unable to finish

1. If you sign the consent form and participate in the discussion but do not complete the survey for whatever reason, you will be given $5 and no bonus. Note that the survey will

not be considered complete if you do not answer the required questions, or if the nickname you report you used in the discussion is invalid.
2. You will receive $2.50 and no bonus if you sign the consent form and make an effort to reach the waiting room, but are not able to reach the discussion in time.

If the study terminates early

You will receive $2.50 and no bonus if you sign the consent form and make an effort to reach the waiting room, but the study has too few people to continue to the discussion stage.

If you do not consent to the study

If you do not consent to participate via the online consent form, you will not be given any compensation or bonus.

Cost

There will be no cost to you if you participate in this study.

Confidentiality

The data captured for the research does not include any personally identifiable information about you. Your IP address will not be captured. We will record your Prolific participant ID.

We will be collecting a record of the discussion between you and the other participants in your group while using our platform. Therefore, we will need to know your "nickname", your discussion username, in order to determine your contributions to the discussion.

By participating in this research, you understand and agree that Carnegie Mellon may be required to disclose your consent form, data and other personally identifiable information as required by law, regulation, subpoena or court order. Otherwise, your confidentiality will be maintained in the following manner:

Your data and consent form will be kept separate. Your consent form will be stored in a secure location on Carnegie Mellon property and will not be disclosed to third parties. By participating, you understand and agree that the data and information gathered during this study may be used by Carnegie Mellon and published and/or disclosed by Carnegie Mellon to others outside of Carnegie Mellon. However, your name, address, contact information and other direct personal identifiers will not be mentioned in any such publication or dissemination of the research data and/or results by Carnegie Mellon. Note that per regulation all research data must be kept for a minimum of 3 years.

Right to Ask Questions & Contact Information

If you have any questions about this study, you should feel free to ask them by contacting the Principal Investigator now through email at slzheng@andrew.cmu.edu. If you have questions later, desire additional information, or wish to withdraw your participation please contact the Principal Investigator through the contact information listed above.

If you have questions pertaining to your rights as a research participant; or to report concerns to this study, you should contact the Office of Research integrity and Compliance at Carnegie Mellon University. Email: irb-review@andrew.cmu.edu. Phone: 412-268-1901 or 412-268-5460.

Voluntary Participation

Your participation in this research is voluntary. You may discontinue participation at any time during the research activity. You may print a copy of this consent form for your records.

I am age 18 or older. ☐ Yes ☐ No
I have read and understand the information above. ☐ Yes ☐ No
I want to participate in this research and continue with the activity. ☐ Yes ☐ No

Appendix D

Instructions

The following instructions, approved by the IRB, was presented to participants just after they signed the consent form.

1. We request you perform the session on a laptop and use either the Firefox or Chrome browser. Please do not use the back button during the course of this session.
2. Please read these instructions carefully. Once you are done, click the "Continue" button.
3. You should see a timer in the upper left corner. This timer indicates how much time you have left before you must be in the waiting room.
4. On the next page, you will follow a tutorial to use the platform.
5. Once you are done with the tutorial, you will be redirected to the waiting room. The "Ready" button will appear around the end of the countdown. **Please do not leave the waiting room during this time.** Once the button appears, please click on it to join the discussion. You will be given a minute to click on the button.
6. If the discussion has enough participants to start, you will be redirected to the discussion. Please input a nickname when requested. **Make sure you remember this nickname so you can record it during the survey.** Please make sure your nickname does NOT have any personal identifiable information, such as your name or date of birth. There should be no way that your nickname can be traced back to you.
7. You will be in the discussion room for 15 minutes with several other people. With this group, you will be solving the "Stranded" problem: **identify and rank the top ten items the group should take if you all were stranded on an island for a long time.**
8. At the end of the discussion, you will be redirected to a survey. It should take approximately 10 minutes. After submitting the survey, you will be redirected to Prolific to indicate the completion of your session and the submission of your response.
9. If the discussion does not have enough participants to start, you will be sent to a page that will let you redirect back to Prolific.
10. If you have any issues during the course of your session, please contact us at slzheng@andrew.cmu.edu.

Payment

Base compensation

If you sign the consent form, participate in the discussion, and complete the survey, you will receive the full payment of $10.

Bonus

After the discussion and in the survey, you will be asked to report the items your group decided to include and say why you've included each item. For each item that *all* members of the group reported was in the ranking with at least a sentence of justification, each member of the group will receive an additional 50 cents in bonus. Maximum $5 bonus from ten such items.

If you are unable to finish

1. If you sign the consent form and participate in the discussion but do not complete the survey for whatever reason, you will be given $5 and no bonus. Note that the survey will not be considered complete if you do not answer the required questions, or if the nickname you report you used in the discussion is invalid.
2. You will receive $2.50 and no bonus if you sign the consent form and make an effort to reach the waiting room, but are not able to reach the discussion in time.

If the study terminates early

You will receive $2.50 and no bonus if you sign the consent form and make an effort to reach the waiting room, but the study has too few people to continue to the discussion stage.

If you do not consent to the study

If you do not consent to participate via the online consent form, you will not be given any compensation or bonus.

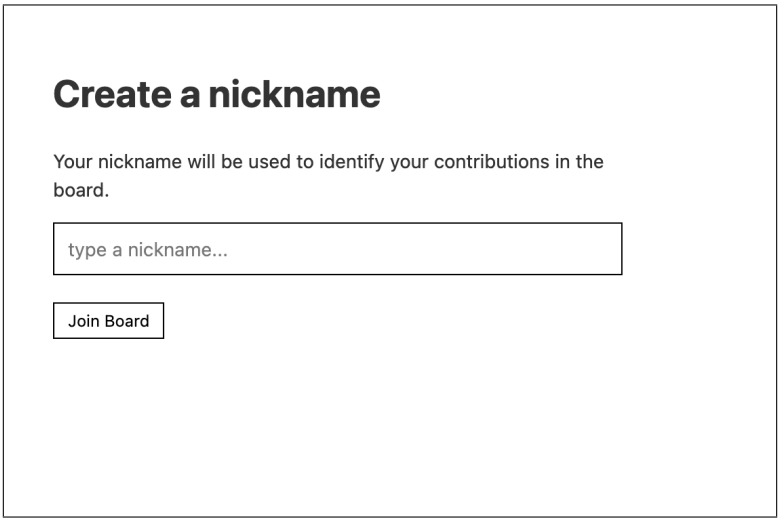
Appendix E

Tutorial

The following tutorials were presented to participants just after they read the instructions.

E.1 Creating a nickname

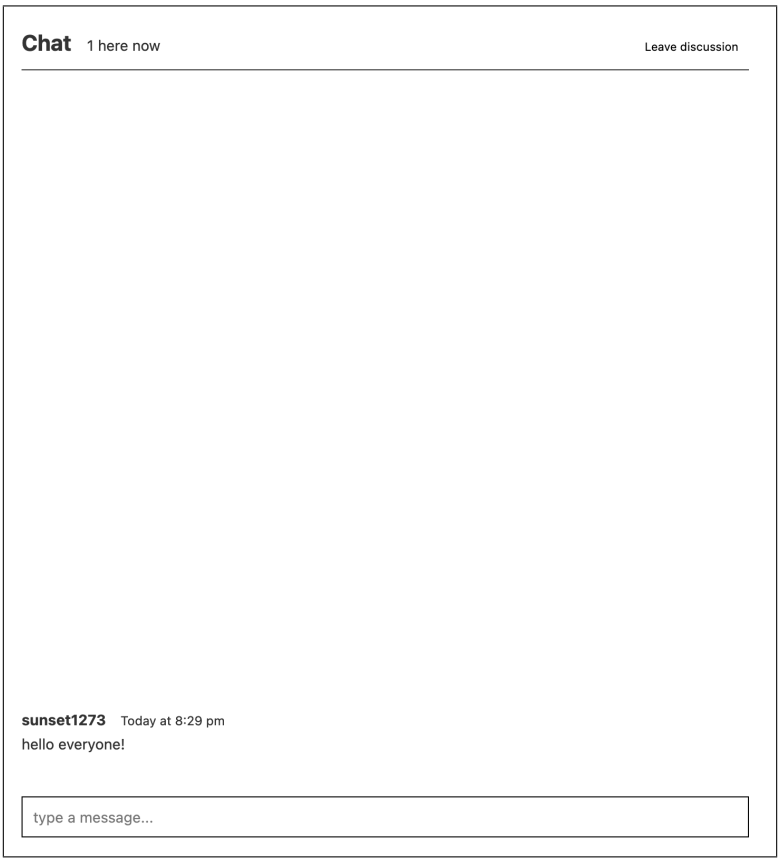
When you join the discussion, you'll be prompted to create a nickname. Keep in mind that other participants in the discussion will see this nickname, so make sure it contains no personally identifiable information.



E.2 The study interface

Once in the study, you'll see an interface that looks like this:

The following was shown for variation 1.



The following was shown for variation 2.

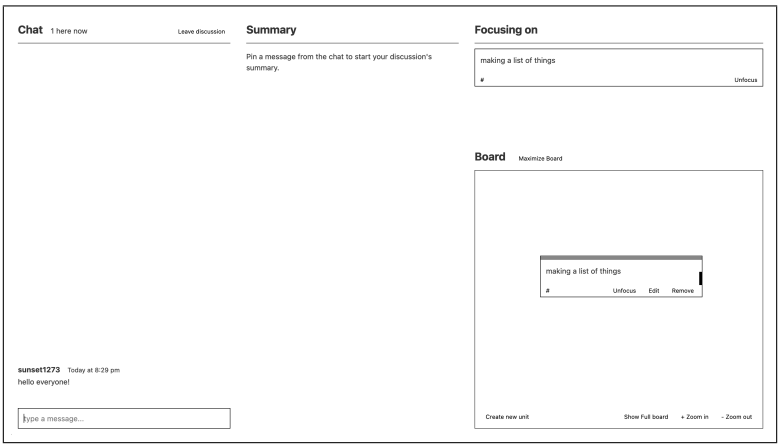
It has two parts:

1. The chat
2. The summary

Each part is described in more detail below.

The following was shown for variation 3.

It has four parts:



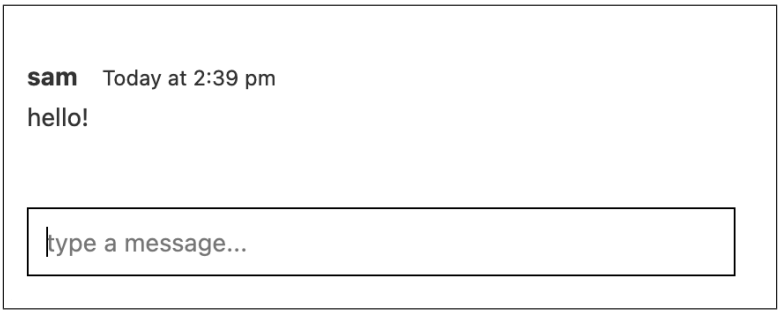
1. The chat
2. The summary
3. The board
4. The focus section

Each part is described in more detail below.

E.3 The Chat

E.3.1 Sending messages

Once in the discussion, you can send messages to other participants by typing in the editor to the lower left of the screen and pressing the enter key.

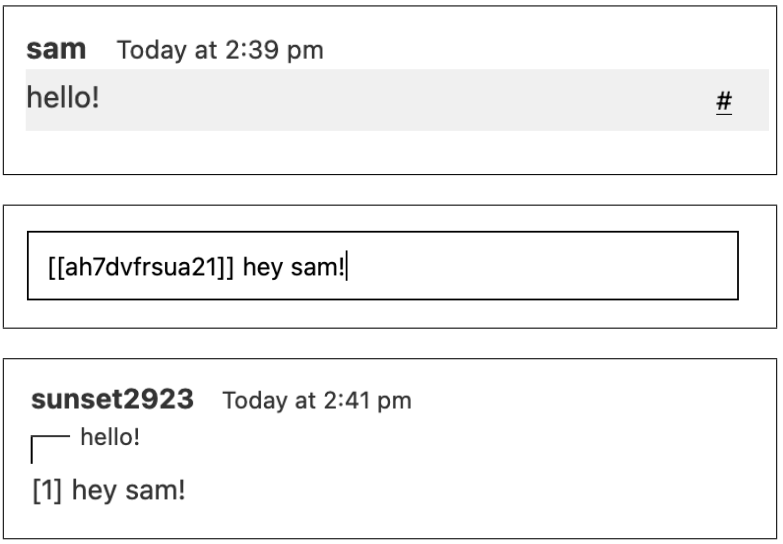


E.3.2 Referencing messages

You can copy a reference to other messages by hovering on a message and clicking on the “”.

Paste the reference in the chat editor, then add any other text.

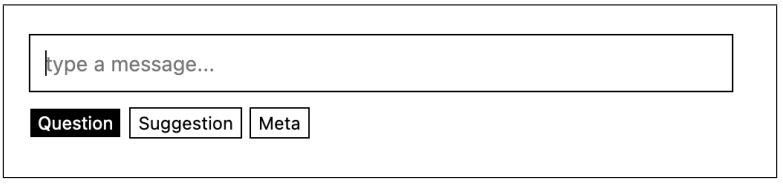
The referenced message will be shown above your message.



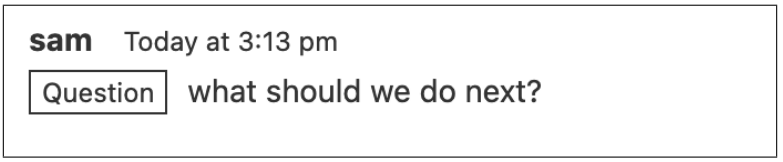
E.3.3 Flairing your message

Added after pilot.

You can add a "Flair" to your message by selecting one or more of the flairs underneath the editor before sending your message.



The flair will be displayed alongside your message in the chat.



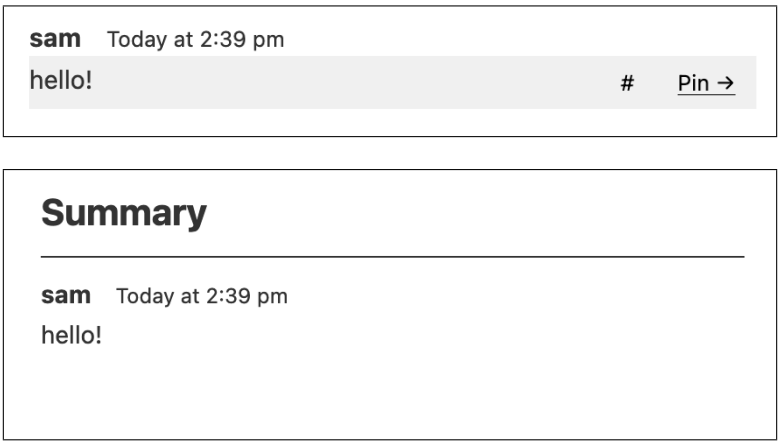
Added for variation 2 after full study 1. In addition, flairing a post automatically pins it to the summary.

E.4 The Summary

Only shown for variation 2 and 3.

E.4.1 Pinning messages

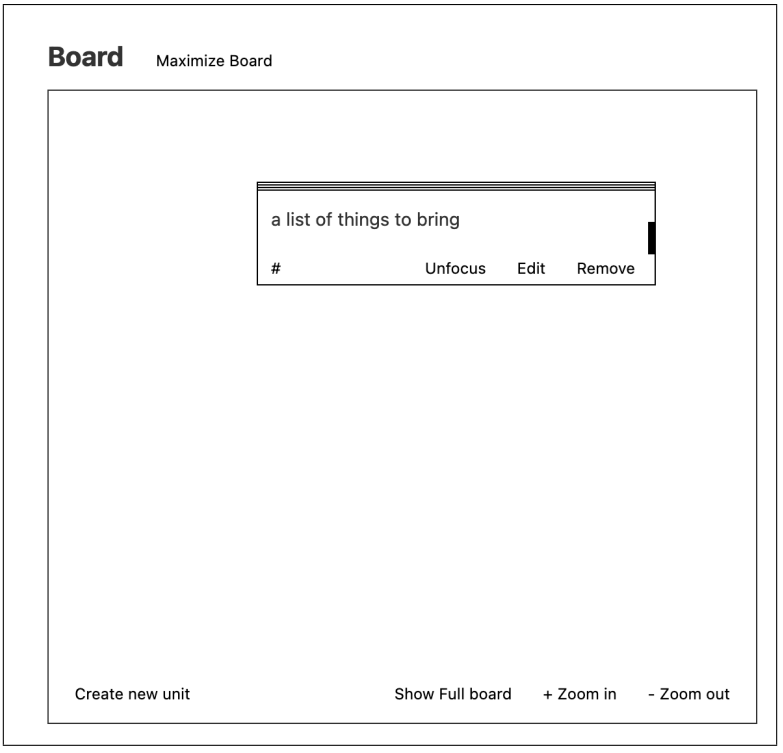
You can add a message to your discussion's "summary" by hovering over the message and clicking "Pin →".



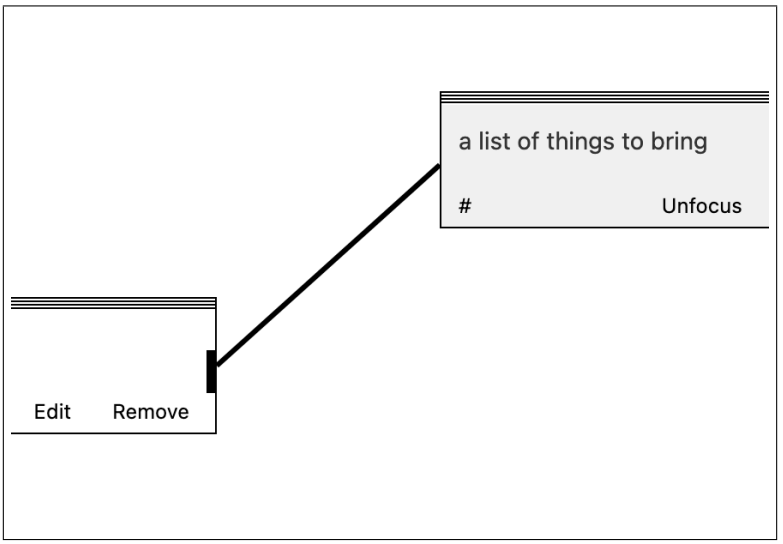
E.5 The Board

Only shown for variation 3.

The board is a two-dimensional space you can place units of text. You can rearrange the units on the board by dragging them from their top bar. You can navigate this board by clicking and dragging, zooming in using a pinch gesture on your trackpad or the scrollwheel on your mouse. You can also use the zoom buttons in the lower right of the screen.



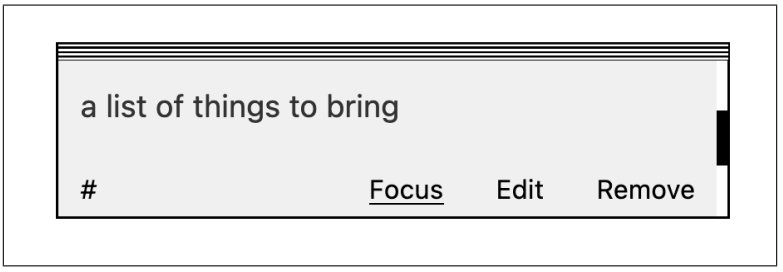
You can create links between units by clicking on the black bar on the right side of a unit and dragging your mouse over another unit.

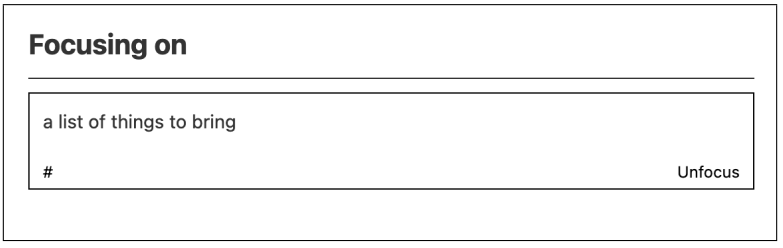


E.6 The Focus Section

Only shown for variation 3.

Pressing the "Focus" button on a unit will add it to the "Focusing on" section of the discussion above the board. Once added, you can remove a unit from the focusing section by clicking "Unfocus".





Appendix F

Survey

The following survey, approved by the IRB, was presented to participants after the discussion to ask about their experience.

Task

You will only be able to continue once you have answered all the required questions.

Short-answer

1. *(required)* Please input the nickname you used during the discussion. **Failure to do this will result in at most $5 compensation instead of the full $10.**
2. *(required)* Please report the items your group decided to include in the final ranking. **Justify each item in about a sentence. Each item justified can earn you 50 cents bonus if everyone else also justifies the item. Max $5 bonus from ten such items.**

Multiple Choice (1 is low, 5 is high)

1. *(required)* Please rate how good you think the group plan is (1-5, 1 being bad, 5 being good).
2. *(required)* Please rate how much you feel you have contributed to the group discussion (1-5, 1 being not at all, 5 being a lot).
3. *(required)* Please rate how much you feel your ideas were heard by the group (1-5, 1 being not at all, 5 being a lot).

Please note down any other observations you would like to share.

Discussion Platform

You will only be able to continue once you have answered all the required questions.

Short-answer

1. *(required)* Please describe what you liked about the platform.

2. *(required)* Please describe what you disliked about the platform.

Multiple Choice (1 is low, 10 is high)

1. *(required)* How easy is it to use the platform (scale of 1-10, 1 being hard, 10 being easy)?

Please note down any other observations you would like to share.

Pilot (given to those in the pilot study)

Short-answer

1. *(required)* Did you use an external aid, such as a notepad, to complete the following question on the previous page: "Please report the items your group decided to include in the final ranking"? If so, please explain what you used it for. If not, just say "no". Note that answering "yes" will not result in a loss of bonus.
2. *(required)* Were there any technical difficulties during the course of your session? If so, please describe what they were. If not, just say "no".
3. *(required)* Was there anything else about the study you felt could be improved? If so, please describe what could be improved and why it should be improved. If not, just say "no".

Please note down any other observations you would like to share.

Bibliography

[1] Gordon Pask. Developments in conversation theory—part 1. *International Journal of Man-Machine Studies*, 13(4):357–411, 1980.

[2] S Buckingham Shum, Karl Aberer, A Schmidt, S Bishop, P Lukowicz, S Anderson, Y Charalabidis, J Domingue, Sara de Freitas, Ian Dunwell, et al. Towards a global participatory platform. *The European Physical Journal Special Topics*, 214(1):109–152, 2012.

[3] DemocracyOS. DemocracyOS. URL https://docs.democracyos.org/. Accessed April 7th, 2021.

[4] Australian Government. yoursay. URL https://yoursay.sa.gov.au/. Accessed April 7th, 2021.

[5] Government Digital Service. Engage with government. URL https://www.gov.uk/government/get-involved#take-part. Accessed April 7th, 2021.

[6] E-Democracy. E-democracy, 1994. URL http://forums.e-democracy.org/about/. Accessed April 7th, 2021.

[7] Matthew Salganik, Karen Levy, and Luke Baker. All our ideas, 2010. URL https://www.allourideas.org/.

[8] Matthew J Salganik and Karen EC Levy. Wiki surveys: Open and quantifiable social data collection. *PloS one*, 10(5):e0123483, 2015.

[9] Synthetron. Synthetron, 2021. URL https://synthetron.com/. Accessed April 7th, 2021.

[10] Kialo. Kialo. URL https://www.kialo.com/tour. Accessed April 7th, 2021.

[11] Debatepedia. URL https://ncdd.org/rc/item/2620/. Accessed April 7th, 2021.

[12] Thoughtgraph Ltd. Debategraph. URL https://debategraph.org/Stream.aspx?nid=61932&vt=ngraph&dc=focus. Accessed April 7th, 2021.

[13] Debate.org. Debate.org. URL https://www.debate.org/. Accessed April 7th, 2021.

[14] Esin Durmus and Claire Cardie. Modeling the factors of user success in online debate. In *The World Wide Web Conference*, pages 2701–2707, 2019.

[15] Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams. Twitter, 2006. URL https://twitter.com/. Accessed April 7th, 2021.

[16] Steve Huffman, Aaron Swartz, Alexis Ohanian, and Advance Publications. Reddit, 2005. URL http://reddit.com/. Accessed April 7th, 2021.

[17] Eugen Rochko. Mastodon, 2020. URL https://joinmastodon.org/. Accessed April 7th, 2021.

[18] Olli Tiainen and Mattermore. Mattermore. URL https://www.mattermore.io/. Accessed April 7th, 2021.

[19] Elliott Cost. Special fish. URL https://special.fish/. Accessed April 7th, 2021.

[20] Facebook. Messenger, .

[21] Stewart Butterfield, Eric Costello, Cal Henderson, Serguei Mourachov, and Slack Technologies. Slack, 2013. URL https://slack.com/. Accessed April 7th, 2021.

[22] Discord Inc. Discord, 2015. URL discord.com. Accessed April 7th, 2021.

[23] Quill. URL https://quill.chat. Accessed April 7th, 2021.

[24] Facebook. Workplace, .

[25] Eric Yuan. Zoom video communications, inc., 2011. URL https://zoom.us/. Accessed April 7th, 2021.

[26] Lalit Mangal. Airmeet, 2019. URL https://www.airmeet.com/. Accessed April 7th, 2021.

[27] Gather Presence Inc. Gather, 2020. URL https://gather.town/. Accessed April 7th, 2021.

[28] MakeSpace. Makespace, 2020. URL https://makespace.fun/. Accessed April 7th, 2021.

[29] Excalidraw. URL https://excalidraw.com/. Accessed April 7th, 2021.

[30] Miro. Miro. URL https://miro.com/. Accessed April 7th, 2021.

[31] Brad Campbell and Joseph M Goodman. Ham: A general-purpose hypertext abstract machine. In *Proceedings of the ACM conference on Hypertext*, pages 21–32, 1987.

[32] Donald L McCracken and Robert M Akscyn. Experience with the zog human-computer interface system. *International Journal of Man-Machine Studies*, 21(4):293–310, 1984.

[33] Robert M Akscyn, Donald L McCracken, and Elise A Yoder. Kms: A distributed hypermedia system for managing knowledge in organizations. *Communications of the ACM*, 31 (7):820–835, 1988.

[34] G Halasz, Frank. Reflections on notecards: Seven issues for the next generation of hypermedia systems. *Communications of the ACM*, 31(7):836–852, 1988.

[35] Mark Klein. Achieving collective intelligence via large-scale on-line argumentation. In *Second International Conference on Internet and Web Applications and Services (ICIW'07)*, pages 58–58. IEEE, 2007.

[36] Luca Iandoli, Mark Klein, and Giuseppe Zollo. Enabling on-line deliberation and collective decision-making through large-scale argumentation: a new approach to the design of an internet-based mass collaboration platform. *International Journal of Decision Support System Technology (IJDSST)*, 1(1):69–92, 2009.

[37] Mark Klein and Luca Iandoli. Supporting collaborative deliberation using a large-scale argumentation system, 2008.

[38] Ali Gürkan, Luca Iandoli, Mark Klein, and Giuseppe Zollo. Mediating debate through on-line large-scale argumentation: Evidence from the field. *Information Sciences*, 180(19): 3686–3702, 2010.

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

[40] CogNexus Institute. Compendium, 2010. URL https://cognexus.org/download_compendium.htm. Accessed April 7th, 2021.

[41] Tobias Krauthoff, Christian Meter, Michael Baurmann, Gregor Betz, and Martin Mauve. D-bas-a dialog-based online argumentation system. In *COMMA*, pages 325–336, 2018.

[42] SRI International. Structured evidential argumentation system, 2000. URL http://www.ai.sri.com/~seas/. Accessed April 7th, 2021.

[43] Joshua Introne, Robert Laubacher, Gary Olson, and Thomas Malone. The climate co-lab: Large scale model-based collaborative planning. In *2011 international conference on collaboration technologies and systems (CTS)*, pages 40–47. IEEE, 2011.

[44] Chris Janiszewski. The influence of display characteristics on visual exploratory search behavior. *Journal of consumer research*, 25(3):290–301, 1998.

[45] Pirijan Keth. Kinopio. URL https://kinopio.club/. Accessed April 7th, 2021.

[46] Inc. Knowledge Building Concepts. Knowledge forum. URL https://www.knowledgeforum.com/Kforum/products.htm. Accessed April 7th, 2021.

[47] Inc. WorkFlowy. Workflowy, 2010. URL https://workflowy.com/. Accessed April 7th, 2021.

[48] Dynalist. Dynalist. URL https://dynalist.io/. Accessed April 7th, 2021.

[49] Jimmy Wales, Larry Sanger, and Wikimedia Foundation. Wikipedia, 2001. URL wikipedia.org. Accessed April 7th, 2021.

[50] Google. Google docs, 2016. URL www.google.com/docs/about/. Accessed April 7th, 2021.

[51] Ness Labs. Roam research. URL https://roamresearch.com/#/app/help/page/dZ72V0Ig6. Accessed April 7th, 2021.

[52] Notion Labs Inc. Notion, 2016. URL https://www.notion.so/. Accessed April 7th, 2021.

[53] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *International Conference on Machine Learning*, pages 11328–11339. PMLR, 2020.

[54] Amy Zhang and David R. Karger. Wikum, 2017. URL http://wikum.csail.mit.edu/. Accessed April 7th, 2021.

[55] Amy X Zhang, Lea Verou, and David Karger. Wikum: Bridging discussion forums and wikis using recursive summarization. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 2082–2096, 2017.

[56] Centre for Science and The Netherlands Technology Studies, Leiden University. Vosviewer. URL https://www.vosviewer.com/. Accessed April 7th, 2021.

[57] Robert E Horn. Infrastructure for navigating interdisciplinary debates: Critical decisions for representing argumentation. In *Visualizing argumentation*, pages 165–184. Springer, 2003.

[58] Neal Altman, Kathleen M. Carley, and Jeffrey Reminga. ORA User's Guide 2020. Technical report, Carnegie Mellon University, School of Computer Science, Institute for Software Research, Pittsburgh, Pennsylvania, 2020. CMU-ISR-20-110.

[59] Kathleen M Carley. ORA: A Toolkit for Dynamic Network Analysis and Visualization. *Encyclopedia of Social Network Analysis and Mining*, 2014.

[60] W Ben Towne and James D Herbsleb. Design considerations for online deliberation systems. *Journal of Information Technology & Politics*, 9(1):97–115, 2012.

[61] Jim Macnamara. Beyond voice: audience-making and the work and architecture of listening as new media literacies. *Continuum*, 27(1):160–175, 2013.

[62] txti. URL https://txti.es/. Accessed April 7th, 2021.

[63] Guillermo Rauch and Automattic. Socket.io, 2021. URL https://socket.io/.

[64] Willy Tarreau. HAProxy, 2021. URL https://www.haproxy.org/.

[65] Salvatore Sanfilippo and Redis Labs. Redis, 2021. URL https://redis.io/.

[66] MongoDB, Inc. Mongodb, 2021. URL https://www.mongodb.com/.

[67] Solomon Hykes and Inc. Docker. Docker, 2021. URL https://www.docker.com/.

[68] Jordan Walke and Facebook. React, 2020. URL https://reactjs.org/.

[69] Dan Abramov and Andrew Clark. Redux, 2019. URL https://redux.js.org/.

[70] Rich Harris. Svelte, 2021. URL https://svelte.dev/.

[71] Rich Harris. Sapper, 2021. URL https://sapper.svelte.dev/.

[72] JSON Schema, 2016. URL https://json-schema.org/.

[73] Guido van Rossum and Python Software Foundation. Python 3.8, 2019. URL https://www.python.org/.

[74] Canonical Ltd. Ubuntu 18.04, 2018. URL https://ubuntu.com/.

[75] Ryan Dahl and OpenJS Foundation. Node.js, 2021. URL https://nodejs.org/en/.