The Impact of Instant Messaging Features on a Collaboration Platform's Usability

Erik Gilbertsson and Julius Malmström

DEPARTMENT OF DESIGN SCIENCES
FACULTY OF ENGINEERING LTH | LUND UNIVERSITY
2023

MASTER THESIS

TELAVOX

The Impact of Instant Messaging Features on a Collaboration Platform's Usability

Erik Gilbertsson
er5402gi-s@student.lu.se

Julius Malmström
ju1217ma-s@student.lu.se

January 25, 2023

Master thesis work carried out at Telavox AB.

Supervisors: Günter Alce, gunter.alce@design.lth.se
Martin Larsson, martin.larsson@telavox.se

Examiner: Joakim Eriksson, joakim.eriksson@design.lth.se

The Impact of Instant Messaging Features on a Collaboration Platform's Usability

Published by
Department Design Sciences
Faculty of Engineering LTH, Lund University
P.O Box 118, SE-221 00 Lund, Sweden

Subject: Interaction Design MAMM01
Division: Ergonomics and Aerosol Technology
Supervisor: Günter Alce
Examiner: Joakim Eriksson

Abstract

The use of online collaboration platforms has increased drastically in recent years with the rise of remote work. The focus of this Master thesis was to explore what impact different instant messaging features have on the usability of such collaboration platforms. Specifically, the thesis aimed to answer the following research questions: 1) What instant messaging features are essential for good usability in a collaboration platform? 2) How much can an instant messaging feature improve the perceived usability of a collaboration platform? 3) How are new features received by experienced versus inexperienced users?

Through a user-centered, iterative design process utilizing qualitative methods such as heuristic evaluation, user interviews and user-based tests, a prototype reply feature was developed for an existing collaboration platform. Usability evaluations conducted on both the original and updated versions of the platform showed that the usability was significantly increased by the addition of the new feature.

The findings indicate that certain instant messaging features can significantly impact the usability of a collaboration platform, and that features tackling issues with organization, structure, and findability are prioritized by users. Lastly, lack of experience in a platform does not seem to affect how a user perceives the usability of the platform.

Keywords: Collaboration Platforms, Instant Messaging, Chat, User-Centered Design, Usability

Sammanfattning

Användandet av samarbetsplattformer har ökat drastiskt under senare år i och med att distansarbete blivit allt mer vanligt. Målet med den här masteruppsatsen var att utforska vilken inverkan olika chattfunktioner har på användbarheten av sådana samarbetsplattformer. Mer specifikt syftade uppsatsen till att svara på följande forskningsfrågor: 1) Vilka chattfunktioner är avgörande för god användbarhet i en samarbetsplattform? 2) Hur mycket kan en ny chattfunktion förbättra den uppfattade användbarheten av en samarbetsplattform? 3) Hur tas nya funktioner emot av erfarna respektive oerfarna användare?

Genom en användarcentrerad, iterativ designprocess med kvalitativa metoder så som heuristisk utvärdering, användarintervjuer och användarbaserade tester, utvecklades en prototyp-funktion för att svara på chattmeddelanden i en befintlig samarbetsplattform. Användbarhetsutvärderingar som utfördes på både den ursprungliga och uppdaterade versionen av plattformen visade att användbarheten ökade signifikant efter tillägget av den nya funktionen.

Resultatet visar på att vissa chattfunktioner kan ge en signifikant ökning av användbarheten i en samarbetsplattform och att funktioner som löser problem relaterade till organisering, struktur och sökbarhet prioriteras av användare. Slutligen verkar det inte som att brist på erfarenhet i en plattform påverkar hur en användare uppfattar användbarheten av plattformen.

Nyckelord: Samarbetsplattformer, Direktmeddelanden, Chatt, Användarcentrerad Design, Användbarhet

Acknowledgements

First we would like to express our gratitude to our supervisor Günter Alce at the Department of Design Sciences, LTH, for guiding us and sharing his expertise throughout the design and writing process. We would also like to thank Martin Larsson who welcomed us into Telavox and provided us with ongoing support and encouragement.

Finally, we would like to thank all of the individuals who participated in our user tests and interviews for their time and valuable insights. Without your contributions this Master thesis would not have been possible.

Contents

1 Introduction **4**
- 1.1 Background . . . 4
- 1.2 Related Work . . . 4
- 1.3 The Sustainable Development Goals . . . 5
- 1.4 Purpose and Research Questions . . . 5
- 1.5 Scope and Delimitations . . . 5
- 1.6 Thesis Structure . . . 6

2 Methodology **7**
- 2.1 The Design Process . . . 7
- 2.2 Gathering and Analyzing Data . . . 8
- 2.3 Prototyping . . . 10
- 2.4 Usability Evaluation . . . 11

3 Initial Product Analysis **14**
- 3.1 Literature Review . . . 14
- 3.2 Heuristic Evaluation . . . 15
- 3.3 User Interviews . . . 17
- 3.4 Affinity Diagram Analysis . . . 18

4 Iteration 1: Lo-Fi Prototyping **21**
- 4.1 Idea Generation and Prototyping . . . 21
- 4.2 User Tests . . . 25
- 4.3 Take-aways . . . 28

5 Iteration 2: Mid-Fi Message Reply Prototype **30**
- 5.1 Idea Generation and Prototyping . . . 30
- 5.2 User Tests . . . 33
- 5.3 Take-aways . . . 35

6 Iteration 3: Final Prototype and Evaluation **36**
6.1 Idea Generation and Prototyping 36
6.2 SUS Evaluation 39
6.3 Results 43

7 Discussion **46**
7.1 The Design Process 46
7.2 Answers to Research Questions 49
7.3 Future Work 51

8 Conclusion **52**

Bibliography **53**

Appendix A SUS Questionnaire **56**

Appendix B Interview Schedules **57**
B.1 Interview Schedule (Initial Product Analysis) 57
B.2 Post-test Interview Schedule (Iteration 2) 58

Appendix C Final Prototype **59**

Chapter 1
Introduction

1.1 Background

Demand for online collaboration platforms has increased drastically in the past few years. Microsoft Teams, launched in 2017, surpassed the previous market leader Slack by reaching 19 million monthly active users in 2019 [18]. This number quickly grew to 145 million in 2021 and now 270 million in 2022 [24, 5].

This master thesis is conducted together with Telavox, a company that provides telephony and customer contact solutions together with a business collaboration platform which here on after will be referred to as *the Telavox application*. The Telavox application features a basic instant messaging tool and online video conferencing, both integrated with the Telavox telephony service. The focus of this thesis is on the instant messaging functionality of this application.

In order to compete with the likes of Microsoft and Slack, it is important to understand what features the users expect and how they contribute to the workflow and usability of the platform. The thesis hopefully sheds some light on this area by identifying key features for instant messaging in collaboration platforms and evaluating what impact they can have on the usability of the platform.

1.2 Related Work

Although there are several interesting articles that have explored the usage of different collaboration platforms in software engineering, few have focused on the impact that different instant messaging features can have on the usability of a platform. Stray and Moe [22] discuss the use of the collaboration platform Slack in a software engineering company, mentioning several features and their impact on communication between employees. Their findings highlight some issues present in every-day use cases of the platform. A similar study by Nyktarakis

[7] studied the usage of Microsoft Teams in different agile software development teams and how it enabled them to work remotely.

In a previous master thesis at Telavox, Glas and Christiansson [23] implemented and evaluated an instant messaging search tool for the Telavox application. They presented their findings on how such a feature can be implemented and quantified its success by measuring the perceived usability of the search tool. Although Glas' and Christiansson's paper is similar to this one in its methodology, an important distinction is that this thesis explores several different features and how they can impact the overall perceived usability of the Telavox application.

1.3 The Sustainable Development Goals

This thesis aims to contribute towards goal number 8, "Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all", from the Sustainable Development Goals set by the 2030 Agenda for Sustainable Development [17]. The hope is that by providing new knowledge regarding how to design and implement collaborative software, businesses can provide their employees with tools for a better remote work environment, facilitating productive employment and sustainable economic growth.

1.4 Purpose and Research Questions

The main goal of the thesis is to explore what features are important for improving the usability of a collaboration platform. Furthermore, the thesis investigates how such a feature can be implemented on a modern platform like the one provided by Telavox and how much it affects the perceived usability of the platform.

This paper aims to answer the following research questions:

1. What instant messaging features are essential for good usability in a collaboration platform?

2. How much can an instant messaging feature improve the perceived usability of a collaboration platform?

3. How are new features received by experienced versus inexperienced users?

1.5 Scope and Delimitations

To answer the research questions, the thesis explores different instant messaging features and how they impact the usability of the platform. The work was delimited to the web version of the application, ignoring other versions like the smartphone or desktop application. Due to this, the results do not show how potential changes would look like or be perceived by smartphone or desktop users, but more time could be spent developing and designing prototypes for the web version.

Another delimitation of the project was the test user group. With Telavox's customers being off-limits for testing, the design process was instead centered exclusively around Telavox

employees who use the application in their work, with some being members of the application's development team. The reason for this was that doing research on Telavox's customers can be problematic from a business point of view, as a customer might anticipate a change in the real product.

Finally, with the time limit for the thesis being 20 weeks, it was decided to delimit the prototyping to not include a working high-fidelity implementation in the real application in order to allow more time for individual test sessions and interviewing.

1.6 Thesis Structure

The Master thesis is structured into eight chapters. This first chapter goes over necessary background information, related works, the purpose and research questions of the thesis, as well as how the thesis contributes to the field and to the Sustainable Development Goals. Finally, the scope and delimitation of the work is also presented.

The second chapter introduces the reader to the methodology of the thesis by presenting the design process and methods used for gathering and analyzing data, building and testing prototypes, and evaluating perceived usability in said prototypes.

Chapters three, four, five and six contain the entirety of the design process from the initial literature review to the design of the final prototype. The chapters are divided by the different iterations of prototyping and user testing, each containing test results for the respective iteration, with chapter six containing results from the final usability evaluation.

Lastly, chapter seven provides some answers to the research questions by discussing the results from the design process in relation to previous literature. Some limitations in the design process as well as areas of future research are also discussed. A final conclusion to the thesis is then presented in chapter eight.

Chapter 2
Methodology

This chapter gives a brief introduction to the different methods used in the thesis, including the user-centered design process and the concept of usability. Prototyping, heuristic evaluation, user-based testing and the System Usability Scale are detailed following an explanation of methods for gathering and analyzing quantitative and qualitative data.

2.1 The Design Process

2.1.1 User-centered Design

User-centered design can briefly be described as an iterative process where the product design is driven by different forms of user feedback and test results. The goal of user-centered design is to create a product that supports the users in their actual needs instead of focusing on what is possible with the technology utilized [13].

Three central principles for user-centered design are brought by Gould and Lewis [8]. The first principle is to put an early focus on the users and how they use the product. This is best achieved by involving actual users early in the design process by scheduling interviews or observing them as they use the system. The second principle is to obtain empirical measurements of the usability and learnability of a product by conducting user tests during all stages of the design process. This can be done by observing the performance, thought-process and emotional reactions of users as they perform different tasks in prototypes created for the specific purpose of testing. The third and final principle builds on the second and states that one should use an iterative design approach where the results from the user tests reflects on the design decisions taken for future prototypes [8]. An overview of the iterative design process can be seen in figure 2.1.

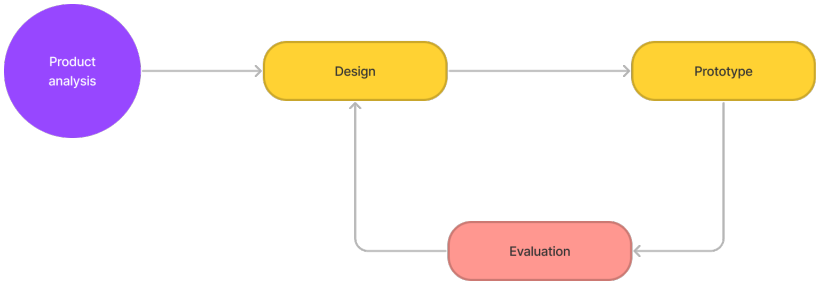

Figure 2.1: The iterative design process.

2.1.2 Usability

Usability can be briefly summarized as how easy and satisfactory a system is to learn and use [13]. For a more in-depth, standardized definition, one can look to ISO standard 9241-11:2018 [10] where three main goals help define how to apply and measure the concept of usability in interaction design.

- **Effectiveness**. Is the system good at doing what it is supposed to do? Can be measured in how many users are able to complete a task.
- **Efficiency**. Is the system too time consuming or complex? Can be measured in how long users take to complete a task.
- **Satisfaction**. Are the users satisfied with the system? Can be measured in frequency of complaints or by rating the system using a scale.

2.2 Gathering and Analyzing Data

2.2.1 Surveys

For the purpose of gathering quantitative data that can be used as basis for a more qualitative gathering of information an online survey with closed questions is used. Surveys or questionnaires allow for easy distribution as it can be distributed both through internet platforms and as physical questionnaires [13]. Unlike a live, physical interview a survey does not require the respondent nor the organizer of the survey to take time and room into account as it can be answered anywhere and anytime [13].

A disadvantage of using a questionnaire is that the information for the recipient is limited to what is written in the form. In a physical interview the person conducting the interview has the possibility to further explain the questions to the interviewee. This leads to the significance of how the questions are formulated in the survey. According to Preece, Rogers and Sharp [13] it is important that the questions have a clear goal and that they are as specific as possible. This minimizes the possibility of misunderstanding for both the respondent and the organizers.

The structure of the survey is important to get valuable answers that help achieve the main goal of the study. Initial questions to get background information about the respondents of the survey such as age and gender can help with the analysis of the answers, as sur-

rounding factors may differ from person to person. The order of which the questions are asked, as well as the length of the survey is also mentioned by Preece, Rogers and Sharp [13] to have an impact on the quality of the survey. Questions can have undesired impact on each other if asked in the wrong order also if they are to long or to short the quality of the data can be questioned.

2.2.2 Interviews

Interviews are commonly used in social research and can take form in many different ways depending on how they are structured [3]. Unstructured and semi-structured interviews are well suited for gathering qualitative data as any issues or miss-understandings can be dealt with immediately and non-verbal factors such as facial expressions and body language can be read from the respondent. Generally speaking, open questions in interviews allows the person being interviewed to uninterruptedly tell their own story which provides for a more nuanced answer [9].

Interview questions can be either open or closed but there is also variants of the two types combined. Closed questions are preferred to obtain quantitative data, open questions are preferred for obtaining qualitative data [9]. Robson and McCartan [3] give some general advice for holding an interview:

- Listen more than you speak.
- Ask straightforward and non-leading questions.
- Avoid industry jargon and interviewer bias.

2.2.3 Quantitative Data Analysis

When analyzing quantitative data it is favourable to start by visualizing the data in tables or using graphical displays like charts or graphs. Summarizing the data by calculating mean or variability can also help in giving an initial understanding of the data [3].

Although more advanced statistical analysis can be used to find interesting correlations, it is not always easy to do correctly without consulting an expert analyst. Simpler analysis is often enough for smaller amounts of data and is even recommended [3].

One example of a method for comparing smaller sample sizes given by two groups in two different conditions is the Mann–Whitney U-test [3, 15]. The test can be used to check the hypothesis that the distributions of two independent sample sets are equal against the alternative hypothesis that one sample set is stochastically larger than the other [15].

2.2.4 Qualitative Data Analysis

Because of its human-centric nature and focus on context and subjectivity, qualitative data analysis can be a good tool when conducting research on people in social settings. Since the data is not quantifiable like with numerical values, one can no longer rely on statistical analysis but instead has to turn to approaches like thematic coding and categorization [3].

A common method for this type of analysis is the affinity diagram. Here the goal is to organise data by looking for connections between individual ideas and visualizing these in a diagram. The affinity diagram is gradually built by categorizing one piece of data at a time and letting new categories and themes emerge during the process. The resulting categorization can then shed light on reoccurring problems and behaviours among the users which can be used to facilitate the design process [12]. Lazar et al. calls this emergent coding and describes it as a process where the researchers find interesting patterns and behaviours amongst data that can be categorized and then used to develop a theory grounded in that data [11].

Figure 2.2: Example of a finished affinity diagram made using post-it notes.

2.2.5 Brainstorming

Brainstorming is an activity with the purpose of generating ideas around a subject or project, often with some data like user responses and feedback as a basis. Some general guidelines for participants in the brainstorming session are to encourage ideas, be non-judgemental of others' ideas and focus on quantity over quality [25].

2.3 Prototyping

Prototyping grants the ability to test and evaluate design ideas during the design process. A prototype can be anything from a paper-based mock interface to a real software implementation with varying degrees of fidelity. Prototypes closer to a real implementation are designated High-Fidelity (Hi-Fi) while cheaper and simpler prototypes are Low-Fidelity (Lo-Fi) [12].

A good example of a Lo-Fi prototype method useful in software development is called Wizard of Oz. The paper or cardboard interface is in this case also reliant on a "behind-the-scenes" human to respond to the user's actions by adjusting the prototype there-after [12, 21].

2.3.1 Hi-Fi or Lo-Fi?

The purpose of the prototype is important to consider when choosing which type to make. Lo-Fi prototypes are significantly cheaper and less time-consuming, making them a good choice for testing many different design ideas in quick iterations. To allow the exploration of different ideas it is preferable that the prototype be easy to modify and adjust [12]. Hi-Fi prototypes can on the other hand be better suited for testing smaller changes to an existing product [21]. They also have the added benefit of giving more complete functionality and being wholly interactive, making it easier to evaluate the user experience as a whole [12].

2.4 Usability Evaluation

To evaluate the effectiveness with which different prototype iterations affect usability one can use a combination of usability evaluation methods. These methods can be divided into categories based on the test environment they are used in. Examples are user-based tests and experiments conducted in a *controlled* environment, field studies conducted with users in their *natural* environment and finally methods involving *user-less* environments like heuristic evaluation or mental walk-throughs [12]. Preece et al. [12] highlight how combining different methods in both controlled and natural test environments can help identify specific usability issues as well as shed light on unforeseen every-day use cases.

2.4.1 Heuristic Evaluation

Heuristic evaluation is a way of identifying lack of usability in a user interface. The process is typically a walk-through of the user interface, where certain principles are brought up and evaluated by experts. The principles that are used can either be well-known universal principles that are used for more than one type of product, or custom principles that have been brought up by doing a competitive analysis on a similar program or interface [19]. The main reason to evaluate the user interface is to communicate the results to designers that later on in the design process can take the evaluation in to consideration when starting new design cycles.

The heuristics that are most commonly used are defined by Nielsen [20] as follows:

1. **Visibility of system status.**
2. **Match between system and real world.**
3. **User control and freedom.**
4. **Consistency and standards.**
5. **Error prevention.**
6. **Recognition rather than recall.**
7. **Flexibility and efficiency of use.**

8. **Aesthetic and minimalist design.**
9. **Help users recognize, diagnose, and recover from errors.**
10. **Help and documentation.**

The heuristic evaluation consists of three stages. First the experts are briefed about the product and the goal of the evaluation. Second, they get time to familiarize themselves with and evaluate the product. Lastly, in the third and last stage, the experts give feedback to the designers of the product. Depending on the level of expertise and complexity of the product the second stage can be repeated several times in order to identify interface components of interest and spot weaknesses in the usability. The experts can also be given a task to carry through in order to steer the experiment in the desired direction [13]. In the resulting list of usability issues each issue should be connected to one of cited heuristics with a clear motivation of how and why it is an issue [19].

2.4.2 User-Based Testing

User-based testing is a form of usability evaluation that can be used both early on and later in the design process. Users are givens tasks to complete using either Lo-Fi or Hi-Fi prototypes, or even the final product itself. Lazar et al. [11] differentiates this as *formative* and *summative* usability testing, where the goal with the former is to be "exploratory and to test early design concepts", while the latter is described as more of a quantitative evaluation of specific design implementations.

When preparing user-based testing there are a number of important questions to consider. How should the users be selected? How many? Where is the test taking place? What tasks are the users going to try? Having the actual user-base be represented when testing is described as "One of the cardinal rules of usability testing" by Dumas et al. [6]. Common factors that should be considered are age, gender, education as well as general technical experience and experience in the system being tested [6, 11]. When it comes to the number of test users, different studies point to around 5 or 7 users being sufficient for most tests, while others say the composition of tasks in the test is what should guide this number. The statistical significance of the results is another factor to consider. In the end time and cost is often the deciding factor meaning one should conduct tests with as many users as is possible within the allotted time and budget [6, 11].

When it comes to where and what, formative testing should preferably be done in the user's natural work space with tasks designed around allowing the user to think aloud, generating qualitative data. Think aloud is a method that encourages the user to speak out loud what he or she is thinking when performing a given task [4]. Summative testing is more forgiving to remote solutions like video or audio calls since the test should be designed around generating quantitative data by measuring the user's performance during and satisfaction after the test [11].

2.4.3 System Usability Scale

Standardized questionnaires can be used to get quantitative data on a user's satisfaction of a system. One such questionnaire that can be used after user-based testing to measure per-

ceived usability is the System Usability Scale (SUS) [14]. The questions in the questionnaire can be found in Appendix A.

Brooke [2] notes how "... just because a particular design feature has proved to be very useful in making one system usable does not necessarily mean that it will do so for another system with a different group of users doing different tasks in other environments." The SUS questionnaire attacks this problem by collecting ten general, subjective measurements of usability which can be applied to any system and thus enables general comparison between systems [2]. This is also true for the comparison of two iterations of the same system, meaning it is possible to measure how individual design choices improve a system's perceived usability [1].

When using SUS the users should be presented with the questionnaire and respond to all ten items immediately after having gone through the different tasks included in the usability test [2]. A SUS test score is calculated using the formula described in [2] which can then be used to compare the usability of systems. A test score above 68 is generally regarded as acceptable [1, 14] while higher scores around 80 to 90 mean the system can be considered to have better or even "truly superior" usability [1].

Although SUS is still considered to be an acceptable method of generating quantitative data after a user test [1, 14, 16], there are some recommendations one should keep in mind when applying the method. McLellan et al. [16] found how previous experience in a system can influence the test score given by a user by as much as 6-15 percent. It is thus recommended that users be asked for their level of experience in the system when evaluating with SUS. Cultural diversity of users and how well they understand the SUS terminology was also found to further influence the test score which underlines the importance of having a good representation of the actual user base amongst the test users. McLellan et al. [16] also recommends that the different scores in the SUS rating system are explained when conducting the evaluation, even if the numerical values are already replaced with adjective scores as recommended by Bangor et al. [1].

Chapter 3
Initial Product Analysis

This chapter describes how the Telavox application and findings in previous research was first analyzed to find possible short-comings and missing features in the application. Quantitative data was gathered through a literature review, heuristic evaluation and user interviews and was then organized in an affinity diagram. The organized data shed light on common themes and problem areas in the application as well as actual feature suggestions which would serve as a foundation for coming up with and deciding on new features.

3.1 Literature Review

3.1.1 Usability Research at Telavox

Glas and Christiansson [23] present an affinity diagram with quantitative data from questions regarding a proposed search feature for the Telavox applications instant messaging tool. The participants of the interview have requests for "Simple design with few mouse clicks", "Good findability, not hidden in a menu" amongst other wishes for a simplistic and clear overview of the chat and the search feature. Further more, when evaluating the final prototype of the search feature using SUS, a sample group of eight Telavox employees gave the prototype an aggregate score of 93 [23].

Other valuable knowledge that can be drawn from [23] is data on the composition of users at Telavox. Glas and Christiansson report that from a survey conducted at the company they concluded that 24 % of the user base is female, 74 % is male and that almost 86 % are between the ages of 20 and 40 years old. Regarding the respondents experience with the Telavox application: a majority of respondents had three or more years of experience and 43 % had taken part in the development of the application.

3.1.2 Software Development Using Microsoft Teams and Slack

In Nyktarakis [7] there are some interesting observations and responses on the use of Microsoft Teams in agile development teams. One common attitude amongst the participants of the study was that they felt they could communicate more easily over chat than traditional communication forms like email or phone calls. Chat communication was perceived as less formal and easier to correct after-the-fact by removing or editing messages. It was also perceived as more immediate than regular face-to-face conversations since everyone could communicate with anyone at a moments notice. One participant specifically mentions emoji reactions being an appreciated way to easily communicate feelings in the chat [7].

When presenting findings around the theme information sharing, Nyktarakis [7] highlights how all participants agree that information sharing had significantly increased following the introduction of Microsoft Teams. Participants responded that communication regarding work mostly took place in different team-wide channels or groups as it was considered important that all team members have the same access to information. The teams participating in the study had several chats for different kinds of information. Team members testified that this had given rise to issues like information sometimes being added to the wrong chat and old information being hard to find in the variety of different chats [7].

Stray and Moe [22] share similar insights into how global software engineering teams use Slack to coordinate their work. They found that Slack facilitates team transparency by allowing the teams to use group channels, though they stress it is also important that the team has a mutual idea of how and when to use these channels. When discussing some mutual principals with a software engineering team, the participants agreed on the following: 1) Communication should preferably take place in group channels instead of private messages, 2) Each team should have a main channel as well narrower channels for more specific issues. 3) Short-lived channels should be used for discussion on specific bugs or features [22].

3.2 Heuristic Evaluation

3.2.1 Approach

A heuristic evaluation of the current Telavox app was conducted to get an initial idea of its overall usability and to find potential weaknesses in the interface. The application was evaluated using Nielsen's ten usability heuristics and the evaluation was divided into two parts as described in section 2.4.1. The first part consisted of getting to know the application's interaction flow and design language by navigating around the interface and trying all features presented. The second part was then spent trying to find specific usability issues using the knowledge gained and the heuristics as a base-point. After having done this individually the resulting usability issues were discussed and then compiled into the following list grouped by which part of the interface they were found in.

3.2.2 Usability Issues

Creating a Chat or Room

- The page for creating a chat could be in violation of the "User control and freedom" heuristic as it does not present a clear way to exit the interaction. The only way to exit is to navigate to another menu or another chat.
- On the right side of the same page there is a menu button which gives an option to create a room. A room is similar to a regular group chat but with options for customization (like naming and making the room private or public) and sending specialized "post" messages. This option can be hard to find, violating the heuristic for "Recognition, rather than recall", and it is not entirely clear what it does because of the naming scheme. Is a room a chat, or is it something else? What is the difference? There is no similar distinction of rooms and chats in competing platforms. This violates "Consistency and standards" and possibly "Help and documentation" as this option could use an explanation before use.

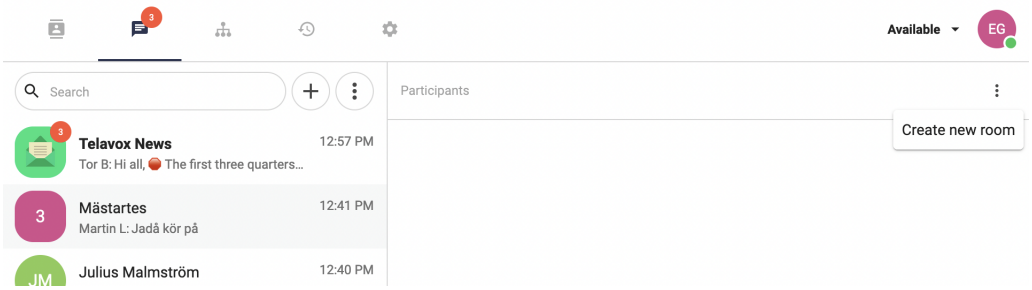

Figure 3.1: The page for creating a new chat or room.

Messaging

- When having sent a chat message, there is no way of editing or deleting the message meaning the interaction does not fulfill the heuristic "Error prevention". There is no way of recouping from a mistake.
- The menu for choosing emojis is problematic regarding "User control and freedom". The only way to close the menu is to click the menu button again which may not always be the first choice as it is not labeled as an exit, possibly trapping users in this interaction.
- Uploading a file in the chat input pane is followed by the filename showing up above the chat input in small grey text. This could possibly be mistaken as the file still being uploaded, as there is no way to discern whether the file is ready to be sent or not, violating "Visibility of system status".
- In a room the user has the ability to "Create a post". This is not available in a normal chat. This is not in accordance with "Recognition rather than recall" as rooms and chats are listed together and can sometimes be indiscernible from each other, meaning

the user has to remember which is which and when creating posts will be available or not.

Figure 3.2: The message input field with an attached file upload.

Chat Options

- Adding new members to a chat does not actually add the member to the chat, it instead creates a new chat with the new member added. This might not be what a user will intuitively expect when using this option as it is described as *adding* the member. This violates the "Match between system and the real world" heuristic.
- In the options menu there are small image cards representing the history of sent images in the chat or room. While this may be an efficient shortcut for experienced users, it might be a bit to minimalist in its design as it is not entirely clear what the images actually represent. As it is not hidden from inexperienced users it is possibly in violation with "Flexibility and efficiency of use".
- Leaving a chat is only possible in public rooms. If a user is added to a private room or a new chat they can only choose to mute or archive these as there is no option to leave. Although this might be by design, it is both an internal and external inconsistency that new users will not be aware of and is thus problematic in regards to the heuristic "Consistency and standards". In competing platforms there are no similar cases where users can not leave channels, groups or chats.

3.3 User Interviews

Interviews were held with five Telavox employees (2 females, 3 males), mean age 30, with the goal of gathering more qualitative data that could be analyzed and later used in idea generation. The interviews were conducted semi-structurally in order to allow the possibility of diving deeper into subjects where the interviewee had strong opinions or wishes for specific improvements. The amount of data collected prevents it from being displayed here, so it is instead presented in the following section as part of the qualitative analysis conducted using an affinity diagram.

Questions were prepared before the interview and added to an interview schedule along with a short briefing on the purpose of the interview. The interview consisted of a first set of questions establishing the interviewees background at the company and their experience with different chat applications. The next set of questions then asked the interviewee to compare these applications and explain which they prefer and why. Several follow-up questions were prepared in the case that the interviewee mentioned issues that had been found in the heuristic evaluation or earlier research. For example: one follow-up was how the interviewee

would describe their teams use of the chat in regards to communication and information sharing. How many and what type of chats did they use? What kind of information did they share in the application? Was there any room for improvement in this regard? The entire interview schedule (in Swedish) can be found in Appendix B.1.

3.4 Affinity Diagram Analysis

The data extracted from the user interviews, heuristic evaluation and previous research was put together in an affinity diagram. Affinity diagrams allows for gathering and organizing of data and is explained in section 2.2.4. Post-it notes in three different colors where chosen to represent the different methods used for the data gathering. Blue notes represented observations from previous research, yellow notes were issues found in the heuristic evaluation and pink notes were quotes and ideas from the user interviews. The reason for this color scheme was mainly to get a clear overview of what the different perspectives had in common and in which ways they differed. It also helped identify potential biases in the different sources. The sorting algorithm for the post-it notes were the following:

1. **Pick a Post-it note.**
2. **Put it in the group of notes it correlates the most to.**
3. **If no group is found, make a new group or put the note aside.**
4. **Discuss and make changes to the current state of the diagram.**

This was reiterated until there were no notes left and all groups were agreed upon to not be discussed further. The resulting affinity diagram can be seen in figure 3.3.

Figure 3.3: The final version of the affinity diagram.

3.4.1 Groups and Take-aways

As can be seen in figure 3.3 the resulting diagram consisted of five groups and in the case of organizing messages it was decided to further divide them into subgroups. The reason to this was that the group could be refactored further to clarify different views and perspectives whilst maintaining the structure and relate them to the headline of the group.

Chat Organization

The first group, *Chat Organization*, identified the overall desire of having some sort of structure in the way chats are organized. This was found from data present in both the interview responses and literature research with observations and responses often pointing to users having a large number of different chats making the interface cluttered and hard to navigate. In the interview responses there were multiple suggestions on how this could be improved such as letting the user color code chats or mark certain chats as "favorite".

Group Chats and Rooms

The take away from the second group, *Group chats and rooms*, was that some users inhibit a sense of frustration and confusion on the topic of how a room or a group chat is created and maintained. This notion was found in the data from both the heuristic evaluation and the interviews. One user pointed out that they did not know whether their team used group chats or rooms when communicating, or what the difference was. This idea is also supported

by one of the findings from the heuristic evaluation. Another user mentioned that they saw no advantage in using a group chat over a room and that they found group chats irritating because of how adding members would create new group chats.

Features

The third group, *Features*, consisted entirely of data from the interviews in the form of feature requests. Giving the interviewees freedom in the sense that there would be no obstacles in implementation, they came up with feature requests like integration of external software such as Jira and Git, and the ability to add bots in the chat that would possibly help in daily tasks like notifying when a CI/CD job was ready. One of the most requested features mentioned by four interviewees was a chat search feature. Since this has been investigated in a previous master thesis, as referenced in section 3.1.1, it was decided to scratch this from the possible prototype candidates.

How Chats Are Used

The fourth group, *How chats are used*, shows that the chat function is a central part of communication and collaboration in a business and gives some important insight into the importance of information sharing through team-wide chats. One user mentions the issue of colleagues sometimes only sending questions to one member of a team instead of involving the entire team which would probably yield a faster response. Users also mentioned there being some chats solely used for less important messages such as where to eat lunch or discussion of hobbies, while other chats focused on more business critical messages such as team discussions and announcements. A relevant observation from the literature research was the reoccurring issue of users having issues identifying the correct chat for a question. This can possibly be connected to the second group further highlighting the need for some way to organize chats.

Organize Messages

The last group, *Organize messages* has remarks from all three data gathering methods. Having a separate group for this became clear as almost all of the interviewees had comments about different issues with sending, responding to and editing messages. The main take-away from this is that there exists a need for improved methods for responding to messages. Suggestions from users ranged from implementing a way to react with emojis to the most popular feature request: adding a way to directly reply to a message. Out of the five interviewees, three naturally mentioned replying to messages being a feature they like in other chat applications. The remaining two users were positive to the idea of message replies when asked. Three users specifically mentioned the reply "thread" functionality from Slack as a good example.

Chapter 4
Iteration 1: Lo-Fi Prototyping

The first design iteration describing the creation and testing of three different Lo-Fi prototypes. The prototypes were designed following an initial idea generation based on issues found in the previous chapter and were then tested on five Telavox employees in a semi-structured exploratory testing session. Based on the test results a reply feature was decided to be most promising although results and take-aways are presented for all three.

4.1 Idea Generation and Prototyping

The three main take-aways from the affinity diagram in the previous chapter were: 1) Having both group chats and rooms is confusing and not always useful to some users, 2) Some users have a large number of different chats which can easily make the interface become unorganized and cluttered, 3) Users want some way to reply directly to messages. These problem statements were used as the starting point in a brainstorming session with the goal of generating ideas for three new features that could be implemented as Lo-Fi prototypes and later be tested and evaluated.

The Lo-Fi prototypes were made entirely out of paper with a printed version of the start page of the Telavox application as background (see figure 4.1). By choosing to make these first prototypes in paper it was possible to quickly put ideas into reality. Since this first iteration was meant to be mainly exploratory it was also a good way to allow quick updates to the interface during testing, while also drawing focus towards the functionality and usability of the prototype instead of its aesthetics. Having a printed background of the chat was a way to increase the feel of in which context the different features were implemented and to hopefully make the prototype as a whole more recognizable to experienced users.

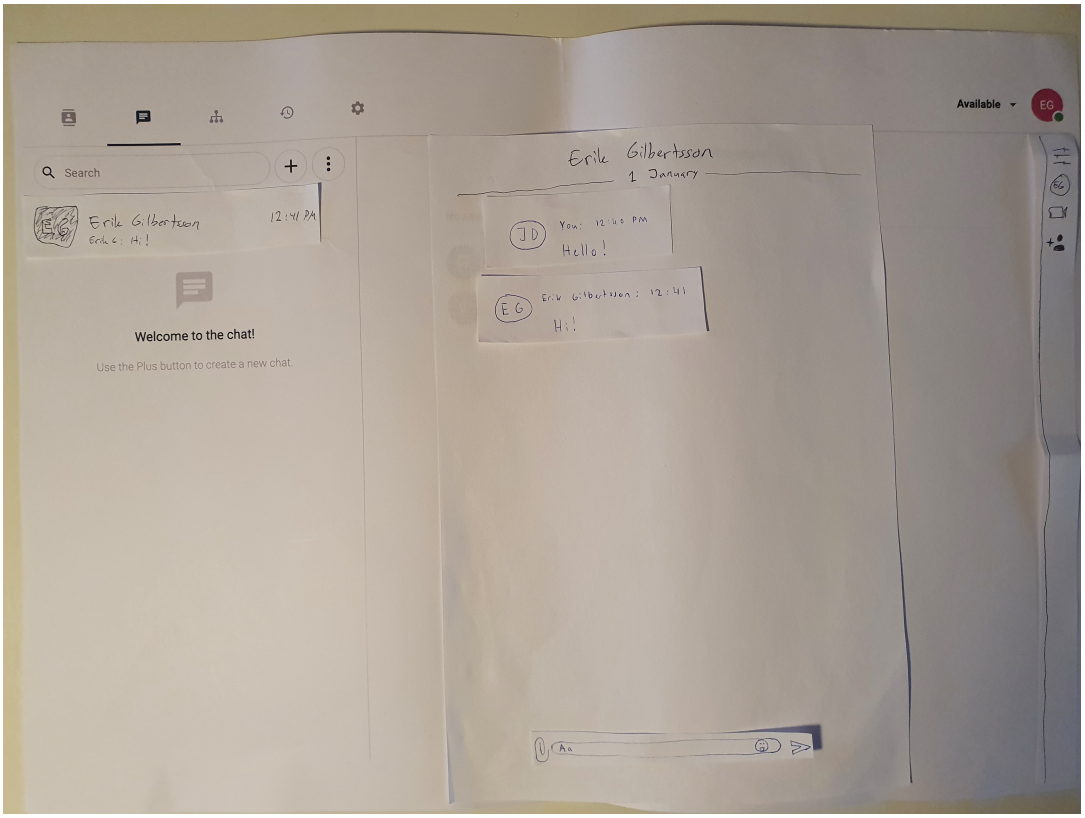

Figure 4.1: Chat interaction in the Lo-Fi prototype.

4.1.1 Feature 1: Simplifying Group Chats/Rooms

Users having trouble with the different concepts of chats and rooms was a hard problem to solve since the solution would need to resonate with both new and experienced users of the application. The current Telavox application has a number of different chat alternatives. The user can create a normal 1-on-1 chat or group chat, or a private or public room. Compared to the normal chat, rooms have some extra functionality like the option to customize the name and icon image. Further more, public rooms are open for anyone to join through a public list. Private rooms require an invite.

The initial idea was to simply merge group chats and private rooms into one entity with all the features of the private room. This would let the users create either a private or a public room for group conversations while keeping the regular chat functionality for 1-on-1 chats.

To further solve the issue of some users finding it confusing with different features for different kinds of chats and rooms it was decided to go further and remove the regular chat concept entirely. This meant 1-on-1 chats would also be private rooms allowing anyone to add new members directly to any chat as well as send posts, regardless of how the chat was created.

One issue that was discussed regarding this approach was how to change the naming scheme. Since all chats would now be rooms, the obvious solution would be to simply call it a room in order to convey that it would have the features commonly associated with a room. In order to communicate that this was also the new regular chat, it was decided to merge these two words and call the new concept a *chatroom*. A chatroom would be used for both

1-on-1 chats and group chats. If made public it would then become a public room like in the current implementation.

When implementing this feature in the paper prototype, it was decided to put the creation of both chatrooms and public rooms in the same menu under the plus-button. Clicking *Create chatroom* would take the user to the old "Create chat" window (although the chat would now be a private room).

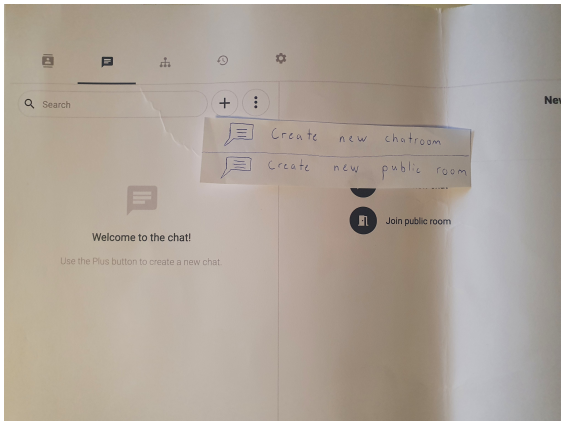

Figure 4.2: Chat or public room creation menu.

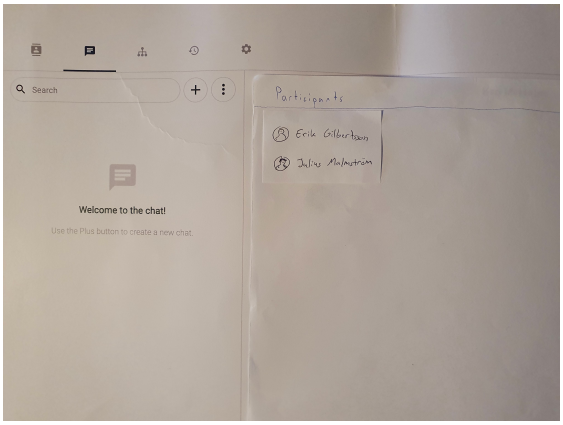

Figure 4.3: Chat creation window.

4.1.2 Feature 2: Chat Organizing Tool

When discussing problem statement two, example solutions like allowing the users to label chats as "favorite" or "pinned" was brought up, but since this issue was found even in the context of applications that already implement these kinds of pinning features, it was decided that the feature should go further. Among the interview responses were wishes for a way to personalize the structure of chats with one user mentioning color coding as a possible solution.

During brainstorming the idea that got the most traction was to implement some sort of folder structure for different chats similar to that of the file system in personal computers, drawing inspiration from the "favorite" drop-down present in many other applications. Users would have the option to create folders, name them and then organize their chats by placing them in whatever folder they like. To keep some similarity to the current system it was also decided that the order of chats within a folder should be chronological. Building on this idea there was also the option to allow the user to switch between the folder view and another completely chronological view in order to not lose out on the advantages of the current implementation.

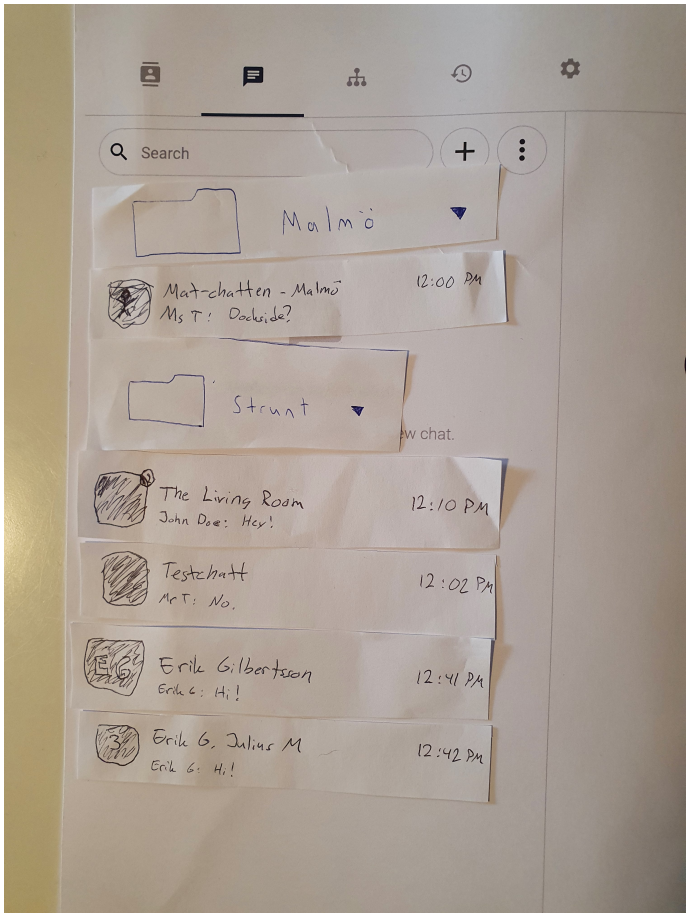

Figure 4.4: Prototype folder view of the chat menu.

4.1.3 Feature 3: Message Replies

There exists several solutions for allowing instant message replies in competing platforms. As can be seen in figure 4.5, replying to a message in Microsoft Teams yields a regular message with the parent message cited above it. Clicking the reply navigates the user to the parent message. In contrast, the team behind Slack has implemented this feature using a separate window for conversation threads (see figure 4.6). In the brainstorming session both of these approaches were mentioned as possible solutions but it was ultimately decided that the Slack approach with conversations threads would be used for the first Lo-Fi prototype. The motivation for this was that three out of five interview respondents had specifically mentioned this as being an enviable feature.

The feature was implemented in the paper prototype as a reply button being shown when the user selects a chat message. When clicked, a reply thread would appear to the right of the chat window allowing the user to reply directly to the message there instead of in the main chat (see figure 4.7).

Figure 4.5: A chat reply in Microsoft Teams.

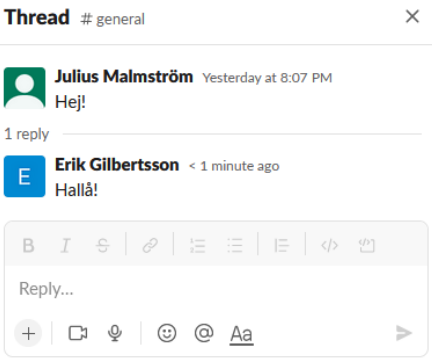

Figure 4.6: A conversation thread in Slack.

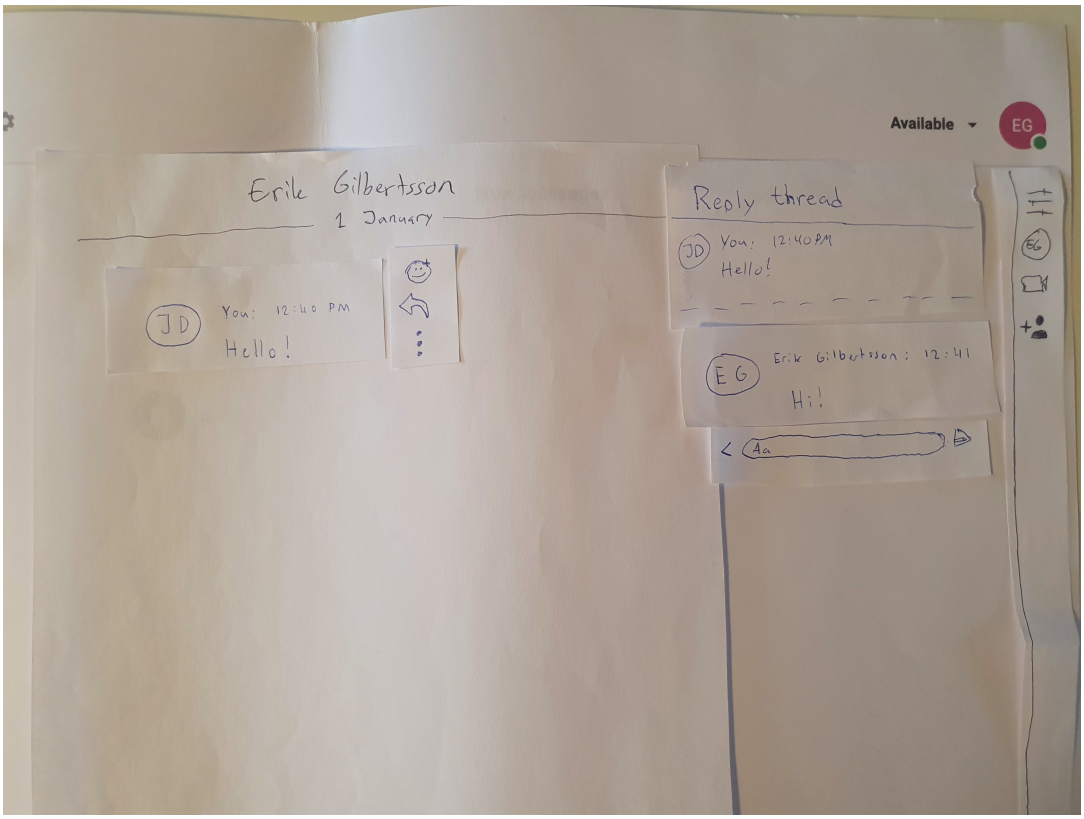

Figure 4.7: Chat interaction using prototype reply thread.

4.2 User Tests

4.2.1 Participants and Test Setup

The user tests for the three Lo-fi prototypes were conducted at the Telavox office in Malmö. A questionnaire was sent out to all Telavox employees with questions on their age, gender, role at Telavox, experience with different collaboration platforms and whether they would like to participate in user testing. Out of ten respondents five were selected to best fit the demographic of Telavox as discussed in section 3.1.1. The average age of the users in this test

was 28.6 years (one female, four males). One user had a non-technical role and four users had technical roles. The average time working at Telavox was 3.3 years.

As mentioned earlier, the nature of this first iteration was mainly exploratory. With this in mind user testing was held in-person at the participants' natural work environment, the Telavox office, with the think-aloud method being applied on different sets of tasks designed around the different prototype features. The test users were to complete these tasks using the prototypes while thinking out loud and explaining their reasoning. Users were also instructed to ask any questions that came to mind during the test.

After each set of tasks a small semi-structured interview would follow which allowed the users to give further feedback on what worked well and what they would like to change. A few questions were prepared for each set, for example: "Can you imagine this feature changing how you use the application? For better or worse?" or "Are there any changes would you like to make?". After the test the user was asked to rank all prototypes ranging from which they like the most to least in order to get quantitative data on which feature was most appreciated.

When testing feature 1 "Simplifying group chats/rooms", tasks were designed to test alternative ways of how a chat is created and how it is maintained when unexpectedly having to add new members. Using the paper prototype test users first performed the tasks listed below with the existing solution and then the new solution.

1. Create a chat with Erik Gilbertsson.
2. Write "Hello!" in the chat.
3. Add Julius Malmström to the chat

For feature 2, the prototype chat organization tool was first shown and explained to the test users. The test users were then given the task of structuring the chat menu however they like. Each user was given free reign and could move around the different folders and structure the chats in the way they would want it to function and look like. Finally for feature 3, message replies, each user was tasked with acting as Erik from the first test and sending a reply to JD using the new reply prototype feature.

4.2.2 Test Results

Feature 1: Simplifying Group Chats/Rooms

The most prominent opinion shared by all five test users was to not show the chat history of a 1-on-1 chat for an added third member. Three users expressed concern about accidentally sharing messages of a private nature. One user even decided to go about adding another member by creating a new group chat from scratch instead of adding them to the private chat since that felt more natural to them.

When presented with the scenario of having three or more members in the chat from the beginning and then adding an additional member, four of the five test users were positive to showing the chat history for new users in the chat, but only if some sort of warning was given to the members first. One person suggested that it should not only be a warning, but an active choice made by the moderator of the chat if history should be visible for new members or not. The one user that was negative to showing all of the chat history was open to the option of showing a configurable amount of messages to the new person being added to the chat.

When asked about their understanding of group chats and rooms, all five test users knew fairly well what the difference between a chat and a room was and how to create them, although an overall positive attitude towards changing the naming and the current placement of buttons to create rooms and chats emerged during the interviews.

One of the more experienced users (over 5 years of experience at Telavox) stated: "There are too many concepts going on in the chat. Decreasing the mind space that is required to use the app would be favorable for the users". Three out of the five users mentioned that a change in naming and placement of buttons would not impact their work significantly, as new chats and rooms are not created often enough. One user argued that while the change would have a positive impact on the usability of the application, it would not have a very noticeable effect on the every day use of most users.

Feature 2: Chat Organizing Tool

One concern expressed by three of the test users was the amount of time and energy the feature would require when committing to actively organize chats with many different settings. Three out of five users thought that it would be more useful if they were instead able to to pin or mark chats and rooms as favorites or muted. Another user stated that they never use the chat menu to navigate between different chats but instead choose to search for whatever chat they need at the moment, making this feature practically useless for them.

Three out of five test users had concerns regarding how new messages would be displayed if everything was divided into folders. In the current solution, the chat/room with the most recent message is displayed at the top of the list. Two users expressed concerns that removing the chronological order would increase the risk of missing urgent messages.

Two test users liked the idea of having separate views for showing the most recent messages in one view and the folder structure were all chats and groups are sorted in another. They both believed that this would bring the most value to the chat. Another user proposed having the two views be displayed together with the chronological view on top of the folder structure view.

A common theme among test users was that having the possibility to ignore or "deprioritize" messages somehow would be more favourable than making folders for commonly used chats. Three out of five users expressed that they would appreciate the folder structure if it meant they could mute or ignore a folder entirely instead of having to individually mute particular chats. One user proposed having the option of muting all chats outside of a folder to more easily focus on one singular project or team.

Feature 3: Message Replies

An overall positive attitude towards the feature could be observed, although there were some points of critique. Two of the test users expressed concern over the lack of findability of new replies in older threads that are no longer visible at the bottom of the chat window. One user mentioned that while it is more clear who the reply is for, it also makes the reply more hidden from other users in the chat.

The users preferred different combinations of having the replies be visible only in the side window and as messages appearing in the original chat window. For example, one user proposed having replies be hidden in the reply thread like in Slack but to also have them show

up in the main chat like in Microsoft Teams. The reason to this was to get both visibility of new answers in ongoing threads as well as a better overview of an entire conversation.

Furthermore, one user argued that some messages should be displayed in the main chat due to the fact that some subjects are meant for the entire group to be seen regardless of ones individual need at that exact moment. Three users felt that having the reply thread be expanded under the message through a drop down menu in the main chat instead of to the side could be an alternative that would suit their needs better. One user with a sales role explained that they would often work with several different windows open on their screen and would therefore prefer if the replies would fit vertically inside the main chat window.

4.3 Take-aways

When asked to rank the three prototypes from which they liked the most to least, the test users unanimously chose the message reply feature as most appreciated. To summarize the results all prototypes were given a score based on the ranking in the different responses. The prototypes were awarded three points when ranked number 1, two points for rank number 2 and one point for rank number 3. The final result can be seen in figure 4.8.

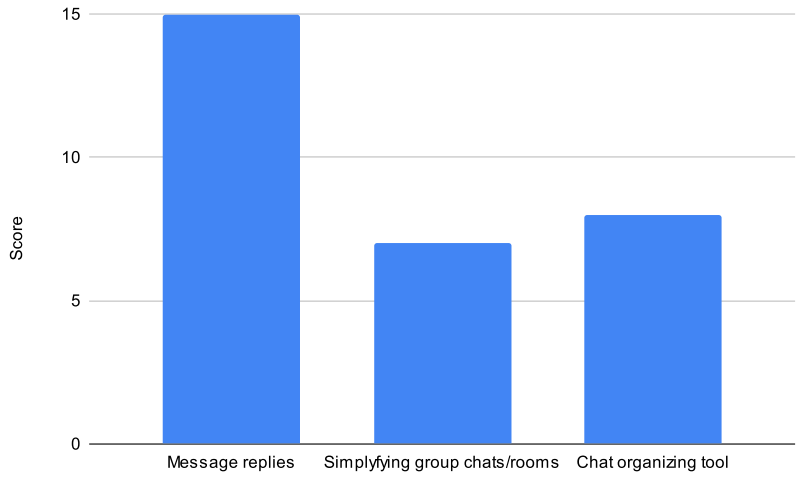

Figure 4.8: Results of the question on which feature the users liked most. Higher score means more desired.

Regarding **Simplifying group chats/rooms**, although all test users did appreciate the change, they still ranked it at second or third place because of the fact that creation of new chats and rooms happen so rarely. An important lesson learnt was that users did not appreciate changing how they add members to a 1-on-1 chat with the reason being that they want to avoid accidentally sharing private chat history with other users.

The main take away from testing the **Chat organizing tool** scenario was that the organizing feature would be appreciated amongst the users if it allows a sufficient amount of different settings to support a large range of use cases. Some users were concerned that the feature could obscure new messages in chats, while other users argued that they wanted the feature to go further by allowing them to completely ignore messages in certain folders. In

addition to this the test users reflected on the fact that not all users would take time to use the organizing tool and get their chats and rooms in order. They therefore argued that the feature would not benefit all users to the same extent as implementing the **Messaging replies** feature might.

The users opinion on **Messaging replies** was the most positive of all three features tested, as can be seen in fig 4.8. A common opinion amongst the test users was that implementing the feature as collapsible columns of replies in the main chat window would be preferable to opening a separate reply thread window. Similar to the test results for the **Chat organizing tool**, some of the test users, although they liked having the ability to directly reply to messages, shared concerns that the feature could possibly obscure important messages outside the reply thread.

Chapter 5
Iteration 2: Mid-Fi Message Reply Prototype

This chapter describes the second iteration of the design process in which the message reply prototype is improved and tested on more users at Telavox. Test results from the previous iteration help guide the prototype design towards a solution which tries to strike a balance in how to structure conversations and reduce message clutter while still keeping messages easy to find for the user.

5.1 Idea Generation and Prototyping

The top rated feature from iteration 1 was the message reply thread which was consequently chosen as the most suitable for continued development. Because of the relatively small nature of the feature as well as the amount of time left in the project plan, it was decided to make this second iteration a Mid-Fi prototype using the interface design application Figma. This would allow the prototype to be more visually appealing and convincing, while still not needing to be entirely functionally or aesthetically complete. Idea generation for possible improvements was based on the two main issues that were identified from user feedback:

1. **Users were concerned important messages would be overlooked if hidden in reply threads.**
2. **Users would prefer to have the reply thread in the main chat window instead of a separate one.**

Regarding the first issue, one user suggested showing replies as new messages in the main chat to help counter the possibility of accidentally hiding important messages for anyone outside of the reply thread. When discussing how to best approach the issue it was decided to also keep in mind one of the root causes for the reply feature being needed in the first place: the high volume of messages in certain chats.

When the need of a reply feature was first discussed in section 3.4, several users mentioned the need of a reply feature in the context of larger group chats where it was hard to reply to a

specific message that had moved up in the chat window due to the high number of new chat messages. With this in mind it was decided that this iteration of the reply feature should somehow tackle the complicated issue of wanting to minimize the number of messages in the chat window while still benefiting from the better findability of showing all messages.

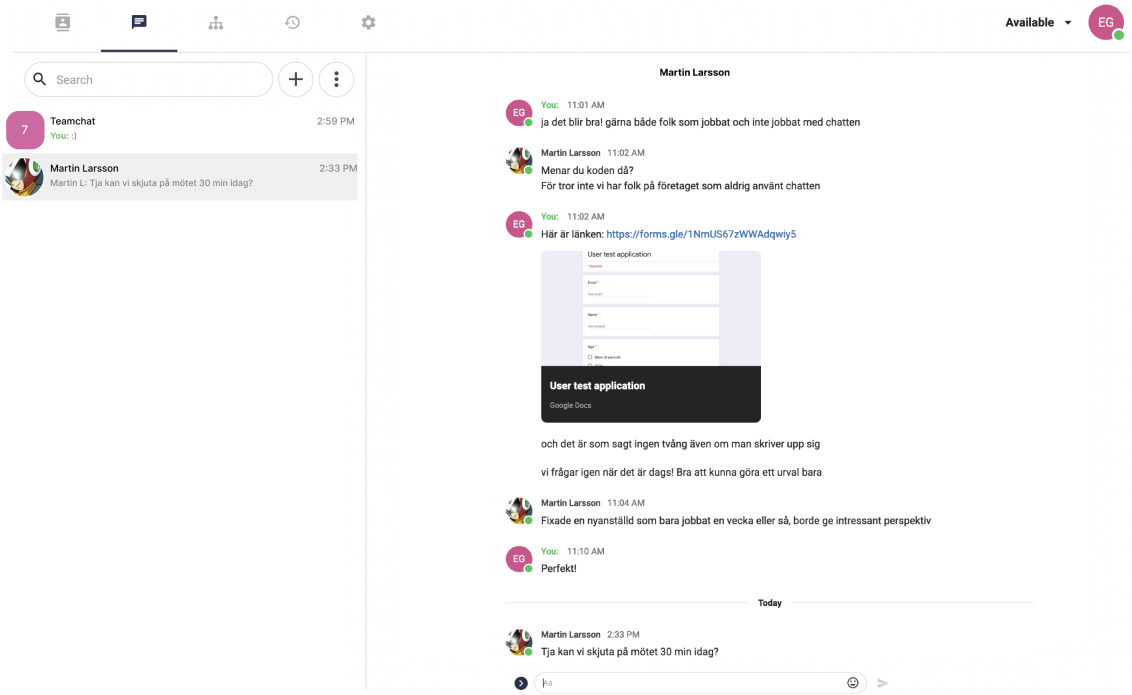

Figure 5.1: Starting view of the Mid-Fi prototype in Figma.

The Mid-Fi prototype can be seen in figure 5.1 and was implemented in Figma using screenshots from the Telavox chat application as a starting point. The implementation is quite similar to the Lo-Fi paper prototype from the previous iteration in how new interactable elements are added on top of a background image of the application. In visual fidelity and functionality the two iterations differ largely as Figma allows easy access to design tools and even automatic animation for certain interactions.

The prototype has three main interactable elements. The first element is the left most menu which can be used to navigate from the chat with Martin to the Teamchat. See figure 5.1. Although this interaction was not strictly necessary for testing the reply feature, it functions as a clear cutoff point between the two tasks used in testing while also making the prototype more convincing.

The second element is the actual reply functionality which is relatively unchanged from the previous iteration in how one chooses to create a reply to a message. Hovering over a message makes a reply-button appear which when clicked will take the user to a new input field for the reply. The difference in this iteration is that the reply thread is now added to the main chat window as a collapsible column of messages under the original message. See figure 5.2 and 5.3.

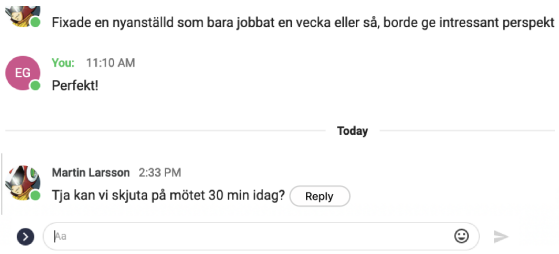

Figure 5.2: The reply button.

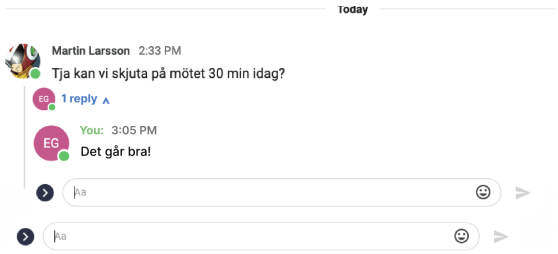

Figure 5.3: A reply and the reply input field.

The third and final interactable element is connected to the issue of how to minimize the number of messages in the chat while still allowing good findability. A compromise was found by showing replies not in the main chat but as pop-up notices in the top part of the chat window. This notice shows the profile picture of the sender and recipient of the reply as well as the actual message. When the pop-up notice is clicked the user is brought to the reply thread in question by scrolling up to and opening the thread as can be seen in figure 5.4 and 5.5. The goal of this design choice was to allow the user to be informed of all messages sent in the chat while still lowering the rate of new messages being added to the main chat.

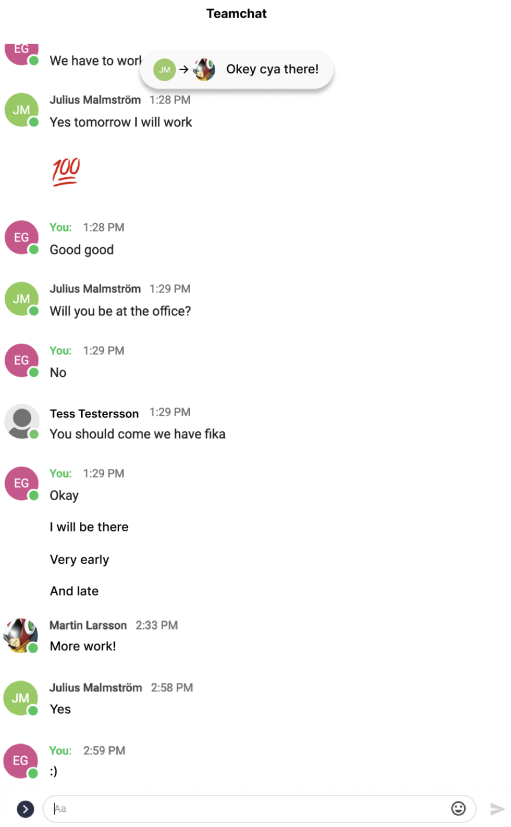

Figure 5.4: The reply pop-up notice

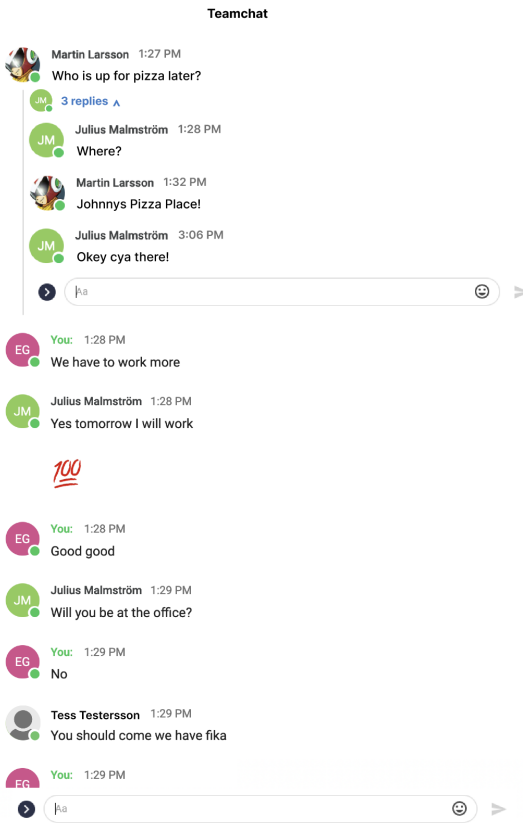

Figure 5.5: Reply thread after clicking pop-up.

5.2 User Tests

5.2.1 Participants and Test Setup

The user tests for the Mid-Fi prototype was conducted at the Telavox office in Malmö. As in previous user tests, age, gender and role at Telavox was gathered to get background information of the users. For this prototype, seven users were chosen from the pool of ten respondents from the previous test session. Two of the participants were recurring testers from the previous iteration. The average age of the users were 34.6 years (two females, five males). Five users had technical roles and two users non-technical roles. The average work experience at Telavox was approximately 3.7 years.

The main goal with the tests was to collect data on how users would react to and use the new reply feature and how it could be improved to fit the different users' needs. Comments regarding the design of specific elements and opinions on how they fit with the current application design were also of interest. The think-aloud method was used throughout the entire test session.

The test consisted of two different tasks: a first easy task and a second harder task. The first task would introduce the user to the new functionality and the second would see how they would use it in a more life-like situation where they would need to find an older reply thread and resume the conversation. After these had been completed, a semi-structured interview was held consisting of eight questions each with 1-2 follow up questions that were asked depending on how nuanced and detailed the respondent was in their initial answer. The full interview schedule can be found in Appendix B.2.

Task 1

The prototype chat for the first task can be seen in figure 5.2 and figure 5.3. The test user was first briefed about the below test scenario and were then given the following task:

- Scenario: You have learnt that a new reply feature has been added to the Telavox application and want to try it out.
- Task: Answer Martin's message about postponing the meeting.

Task 2

The prototype chat for the second task can be seen in figure 5.4 and figure 5.5. The test user was given another test scenario to give them the information required to resume the imaginary conversation in the final, second task:

- Scenario: You receive a new notification from your Teamchat and suddenly remember that you have forgotten to answer Martin's question about who was up for pizza later tonight.
- Task: Tell Martin you'll join pizza night.

In this task the user has two ways to find the reply thread in question. They can either scroll up in the chat and find the thread manually or click the pop-up notification and have the application navigate them to the thread.

5.2.2 Test Results

Task 1

When tasked to reply to Martin's message six out of seven test users immediately moved their mouse over the target message and found the reply button. One user first tried to send a reply through the main chat input but found the reply button when guided. Another user, although they found the reply button, expected the reply to be composed using the main chat input and was surprised about the new chat input.

Following completion of the task, four out of seven users tried collapsing the reply thread without any guidance. The remaining three quickly understood the concept when the clickable blue text was pointed out to them.

All users had an overall positive attitude toward the reply functionality. Three of the users were noticeably excited about the choice of a reply feature.

Task 2

After navigating to the Teamchat in the prototype, four out of the seven test users clicked the pop-up notification and were immediately taken to the reply thread without the need for scrolling. The remaining three users instead chose to scroll up in the chat. Of the users who clicked the notification, two stated before clicking that they expected to be taken to the reply thread while the other two stated they had some notion it was related to Martin's messsage but were unsure exactly what would happen. Of the three users who scrolled up manually to the reply thread, one mentioned they had seen the notification but only understood its purpose after opening the reply thread manually. The other two users had to be shown the notification after the task had been completed.

All users were generally positive to the pop-up notice with some users sharing spontaneous praise and criticism. Two users commented on the notification's visual appearance being nice and one user appreciated not having to scroll. Three users stated that they were unsure how the notification system would function if there were more than one reply thread. One user stated they were concerned it could clutter the chat window.

Interview Responses

When asking the test users if there was anything they would want to change in the prototype, the most common suggestion was around the design of the pop-up notification in task 2. One user suggested that the notification be made more visible by changing the color and increasing its size. Another user mentioned that the purpose of the notification would be clearer if it also showed the parent message. On top of this three test users questioned how the new pop-up notification system would handle multiple reply threads. One user said that the risk of mixing up the notifications generated by different threads would be significant.

Other suggestions were to draw more inspiration from other popular chat applications. One user suggested that all replies be shown in the main chat window like in Microsoft Teams, and another mentioned they would prefer being able to click a reply to navigate to a new view showing the conversation in its entirety like in Apple's iMessage.

When discussing potential upsides and downsides with the reply feature, one scenario that was mentioned by five out of the seven test users was when there are multiple topics

being discussed simultaneously in a chat. Most of these users were of the opinion that the reply feature would help in this regard by allowing them to easily follow the different conversations while at the same time making the chat less cluttered by lowering the amount of new chat messages in the main chat window. One user mentioned that in their experience the problematic "high-volume" chats were typically those with non-work-related topics, making it less of a real issue and more of a nuisance. Another user commented: "I could see this being used by almost everyone on a daily basis."

A potential downside mentioned by three users was the risk that important messages become less visible in reply threads. One user stressed the importance of getting the notification system right to counter this risk. Another user mentioned a search feature for doing look-ups in the chat as a mitigation for the issue.

5.3 Take-aways

An overall positive attitude towards the prototype was observed during the tests and all seven test users answered "Yes" when asked if they would use the feature if implemented. From comments during and after the test it also seems that the users were satisfied with the visual appearance of the prototype.

Similarly to the previous iteration, some users voiced concerns about how the reply feature would impact the findability of the chat. All users were generally positive to the feature with five having practical examples of how it would help them in the scenario of many topics being discussed simultaneously. A few were still slightly hesitant to the idea of hiding the messages in the collapsible reply threads. The concern was that hiding replies would increase the risk of missing important information, although they saw the notification system as a possible solution if improved.

Another issue that was found during the test was how five out of seven test users did not see or did not fully understand the purpose of the pop-up notification in task 2. When shown how it works all users reacted positively, although some users were concerned that with multiple different threads the notification system could become a nuisance and maybe even worsen the ability to keep track of actually important conversations.

When it comes to the use of Figma for this iteration, five out of the seven test users did not struggle in any way when using the Figma prototype to complete the tasks. An explanation of the prototypes limits was sufficient in most cases and when a problem did occur it was enough to go over the limitations again.

Chapter 6
Iteration 3: Final Prototype and Evaluation

This chapter describes the third and final iteration of the design process which includes further improvements to the message reply prototype and a SUS evaluation of the resulting system. The chapter concludes with results from the SUS evaluation conducted on both the current Telavox application and the prototype version.

6.1 Idea Generation and Prototyping

Using Figma to implement and test the reply feature turned out to be a successful approach, it was therefore decided to continue with Figma in this iteration. From the test results three main issues were identified, the first of which remains from iteration one:

1. **Users were concerned important messages would be overlooked if hidden in reply threads.**
2. **Users did not see or fully understand the purpose of the pop-up notification.**
3. **Some users were concerned the pop-up notifications could become annoying.**

Idea generation started with the goal of solving these three issues. It was first concluded that the test users did not think the pop-up notification was a sufficient solution for making replies more visible. It needed to be made more noticeable and at the same time less annoying. There was also need for more improvement in regards to the visibility of information in reply threads.

6.1.1 Improvements

When brainstorming ideas on how to improve the pop-up notification, an obvious improvement was to increase its contrast against the surrounding chat window with more color. An-

other aspect of the issue was that some users did not understand the purpose of the notification. One solution that was discussed was to show the parent message being replied to, although with some users already expressing that they might find the notification annoying, it was decided keep the amount of information to a minimum while still having it be sufficiently self-explanatory.

In order to make it more obvious that the notification is referring to a reply, the names of the sender and recipient of the reply were added. To further decrease the risk of multiple pop-up notifications causing a nuisance for users it was also decided to limit the use of notifications to conversations that the user is engaged in. The resulting design can be seen in figure 6.1

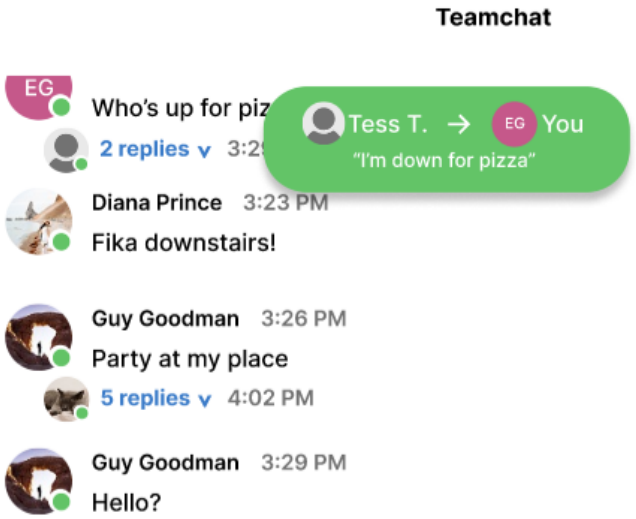

Figure 6.1: The improved pop-up notification.

A recurring theme in all iterations of the design process was the users' concern about message visibility in reply threads. One initial idea of how to go about this issue was to add another complementary system to the pop-up notifications in order to increase the efficiency with which the user can look for information. With the decision to only send notifications for threads that the user has previously participated in, this was deemed as a necessary next step in the design process and it was decided to investigate an alternative method of finding new replies in a chat. This idea took form as a separate view that would filter out all normal messages and only show reply threads sorted chronologically by the latest reply. In this view the latest active threads would show up at the top, with inactive threads being sent further and further down, regardless of how long ago the parent message of each thread was sent. The resulting overview of reply threads can be seen in figure 6.2. The user can navigate to this view by clicking the new icon added to the bottom of the right-side chat menu.

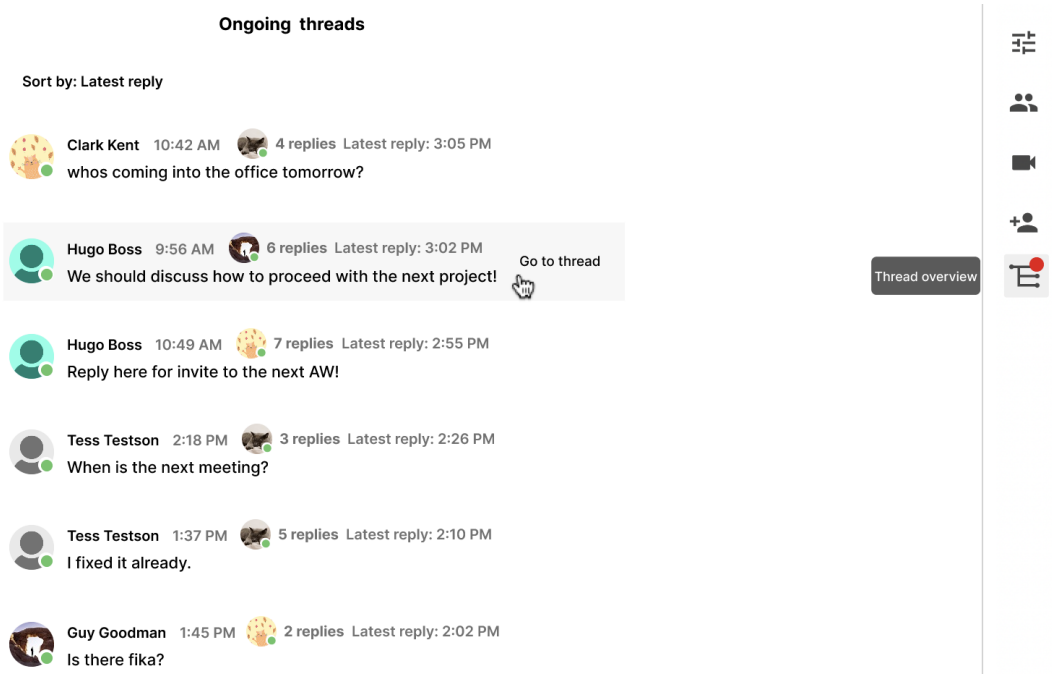

Figure 6.2: Overview of ongoing threads in a group chat.

6.1.2 New Test Scenarios and Tasks

For the purpose of giving the test participants a good representation of how the feature would work in the real Telavox application, the Figma prototype was entirely overhauled with new scenarios. A version of the prototype without the reply feature also had to be created since the SUS score of the reply feature prototype was to be compared to the score of current version of the Telavox application. The Figma prototype *without* the reply feature is referred to as **version 1** and the Figma prototype *with* the reply feature is referred to as **version 2**. It was considered to run the test on the actual Telavox application instead of version 1 of the prototype, but this was decided against in order to avoid the difference in level of fidelity affecting the results.

When deciding on how to design the new test scenarios and tasks, comments from previous tests and interviews were taken into consideration. One theme that was identified in the affinity diagram from section 3.4 was how users want ways to organize messages in a chat. Some users complained about the lack of a reply feature and how it made it harder to track conversations in larger chats. Another common complaint was regarding the issue of not having a clear way of replying to older messages. On top of this the previously mentioned issue of findablity and message visibility was also incorporated into the test. In order to measure if there had been any improvement in these areas, it was decided to incorporate these issues by designing tasks around finding information and replying to older messages.

In total, four new scenarios and tasks were created for the test. Both version 1 and version 2 of the prototype have the same test scenarios, although the wording of some tasks differ slightly because of the lack of a reply feature in version 1. A more in-depth explanation and figures for the new tasks follows in section 6.2.2. Figures showing the entirety of the Figma prototype and the different scenarios and tasks can be found in Appendix C.

6.2 SUS Evaluation

The goal of this final test was to gather quantitative data on the perceived usability of the Telavox application using the SUS questionnaire described in section 2.4.3. The test was conducted on three different groups of users in order to compare the usability of the current and updated versions of the application, as well as to compare the perceived usability between experienced and inexperienced users.

6.2.1 Participants

In total, 33 participants were selected from two populations: people with professional experience using collaboration platforms like Microsoft Teams or Slack (population 1), and employees at Telavox with experience using the Telavox application (population 2). The participants were divided into the following three groups:

- **Group 1:** 11 participants (3 female, 8 male), mean age 29, from population 1.
- **Group 2:** 11 participants (3 female, 8 male), mean age 30, from population 1.
- **Group 3:** 11 participants (3 female, 8 male), mean age 32, from population 2.

The participants in all groups were selected based on their availability and willingness to participate. Participants in groups 1 and 2 were required to have work experience using collaboration platforms other than the Telavox application and participants in group 3 were required to have experience using the Telavox application. The sample set in groups 1 and 2 included fifth-year engineering students at LTH and working professionals from around Sweden while group 3 was strictly made up out of Telavox employees. In order to ensure a diverse range of professional experience, participants from a variety of roles and ages where selected when possible.

6.2.2 Test Setup

The aim of the test was to give the participants an initial impression of the Telavox application which they would afterwards use as a basis for answering the SUS questionnaire. No qualitative data was collected during the test which meant the tests could be carried out remotely over individual video calls with each participant. Group 1 was given version 1 of the prototype and groups 2 and 3 were given version 2.

As with the previous test in Figma, all participants were first briefed about the purpose of the test and what limitations were present in the prototype. The participants without experience in the Telavox application, groups 1 and 2, were given a short briefing on the application as a whole with group 2 also being informed about the new reply feature. Both groups were told they were testing one version of the application with no particular mention as to what was different with their version. Group 3, the Telavox employees, were simply briefed about the reply feature being the new addition.

After a participant had completed all tasks they were immediately asked to answer the SUS questionnaire, following a short explanation of each score on the questionnaire's Likert scale.

Task 1

The first task had the purpose of introducing the test participant to how replying to messages works in their respective version of the prototype. In version 1, seen in figure 6.3, the reply is sent as a regular message from the bottom chat input field. In version 2, seen in figure 6.4, the reply is added to a reply thread under the parent message. Each participant was given the following scenario and task:

- Scenario: You have received three messages from Martin in quick succession.
- Task: Reply to Martin's first message about postponing the meeting.

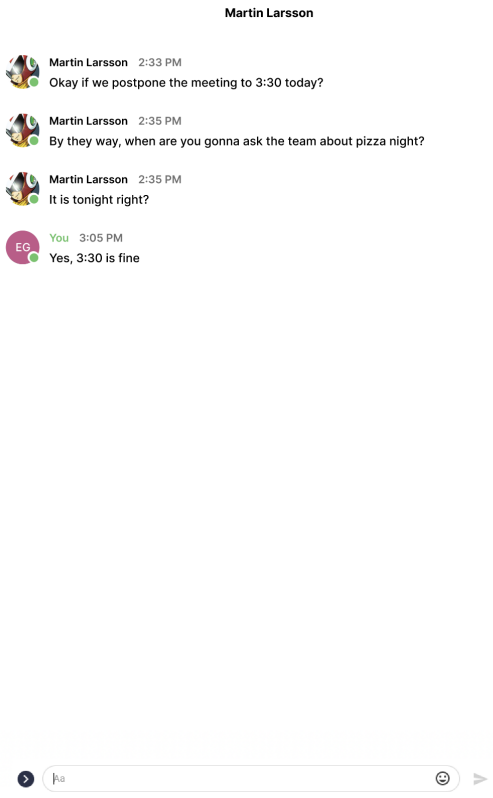

Figure 6.3: Task 1 in version 1.

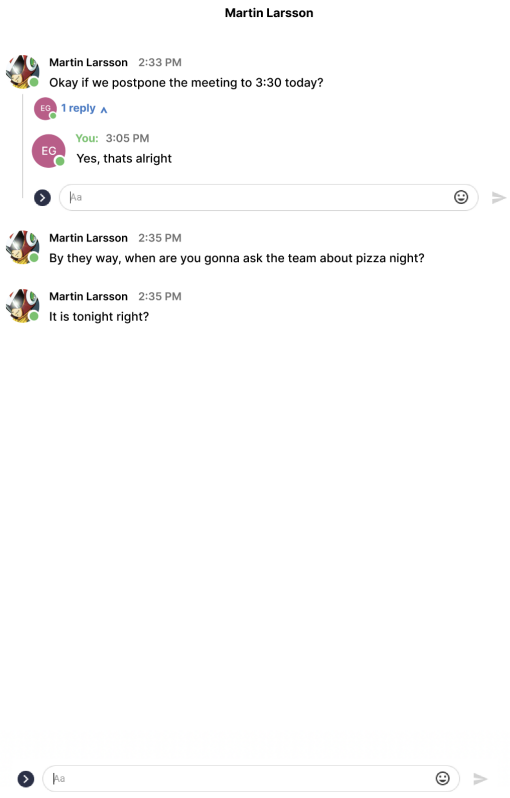

Figure 6.4: Task 1 in version 1.

Task 2

Task 2 is the first part of a two-part scenario used to test how the user can leave a message and then come back to find the conversation later. In this task, there is little difference between the two versions seen in figure 6.5 and 6.6. In both versions participants leave the message using the bottom chat input field with the following instructions:

- Scenario: Martin's second message mentions that you should ask the rest of the team if they are coming to the pizza night.
- Task: Navigate to the "Teamchat" and ask the team who is coming to pizza night.

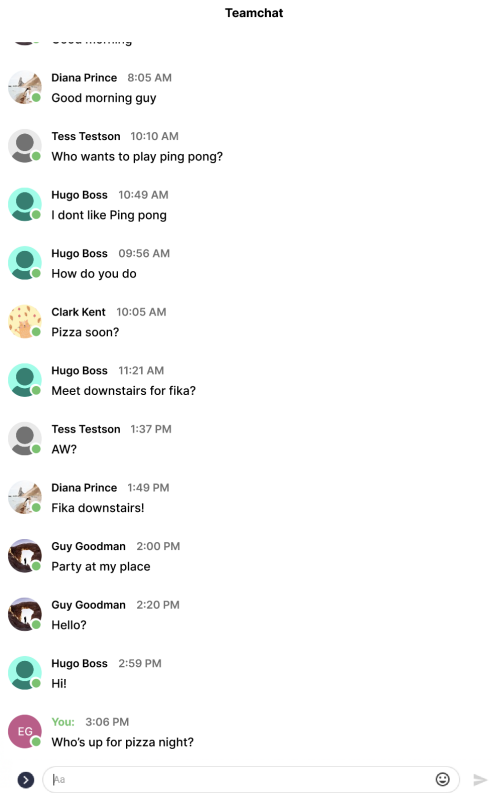

Figure 6.5: Task 2 in version 1.

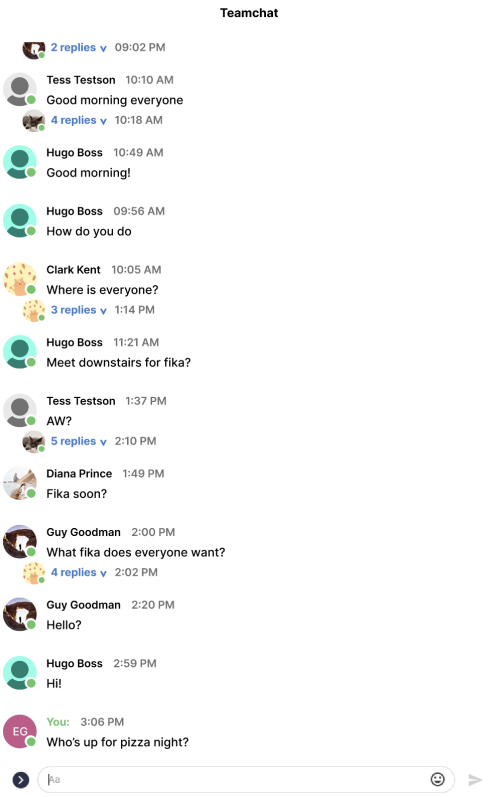

Figure 6.6: Task 2 in version 2.

Task 3

In this task, the participants were to find information in the form of an older conversation in another group chat. The main difference between the two version is that in version 2, seen in figure 6.8, the messages in the reply thread are spread out during the day to showcase how the reply feature keeps the replies as one cohesive conversation. In version 1 seen in figure 6.7, the conversation instead takes place directly following the parent message in order to not make the task unnecessarily hard for the test participants.

Version 2 of the prototype also has two different ways to find the conversation. The participant can either scroll through the chat manually (like version 1) or click the "Thread overview" button in the right-side menu to get to the new reply thread view (see figure 6.2).

The participants were given the following scenario and task:

- Scenario: In "Other teamchat" there is a conversation that you need to read regarding how your team will proceed with the next project.
- Task: Navigate to "Other teamchat" and find the conversation about the project.

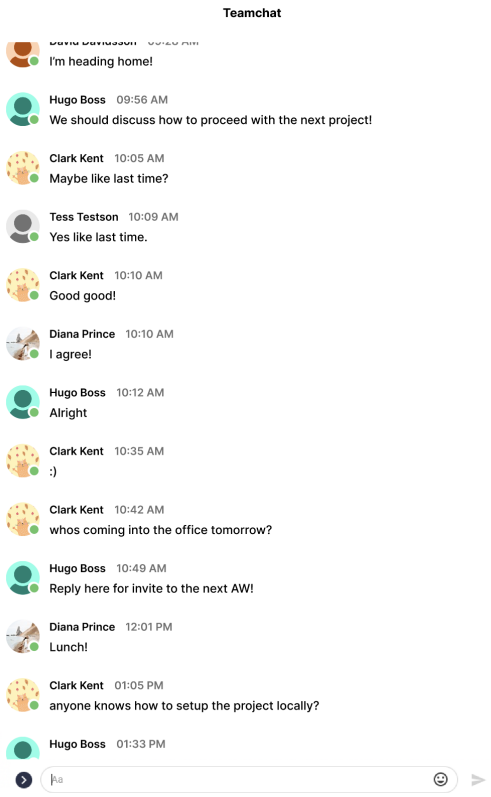

Figure 6.7: Task 3 in version 1.

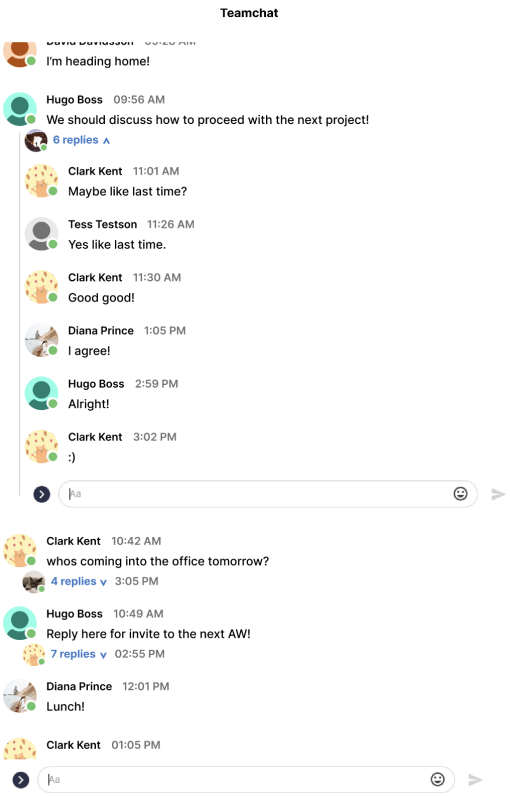

Figure 6.8: Task 3 in version 2.

Task 4

Task 4 is the final task and second part of the scenario set up in task 2. The aim of the task is to again test how easily the user can find information in the chat, but this time in the form of replies to the user's own question.

In version 1 of this task, it was decided to add a competing conversation between replies in order to showcase how present users of the application normally add context to their reply by clearly stating what or who they are replying to (see figure 6.9). Participants with version 2 of the prototype can use the pop-up notification (see figure 6.1) or the "Thread overview" as alternative ways to find the conversation. The task is setup with the following instructions:

- Scenario: We skip ahead an hour in time. You come back from your meeting with Martin and see that there are 22 unread messages in "Teamchat".
- Task: Navigate to "Teamchat" and find out if anyone has replied to your pizza night invitation.

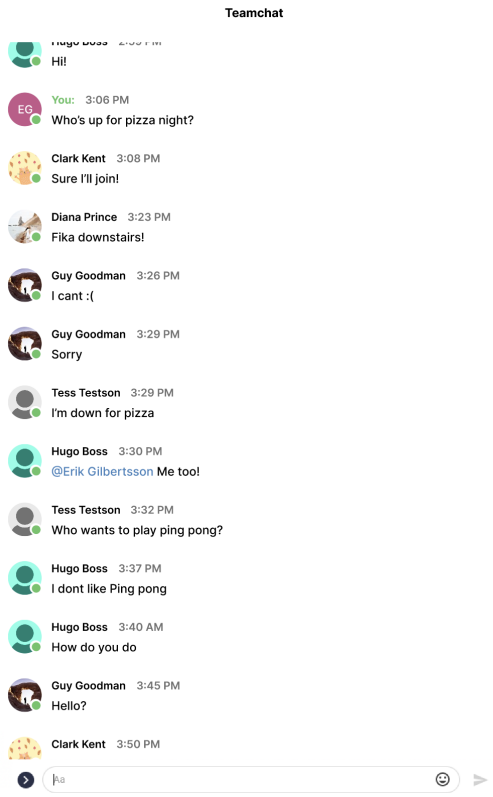

Figure 6.9: Task 4 in version 1.

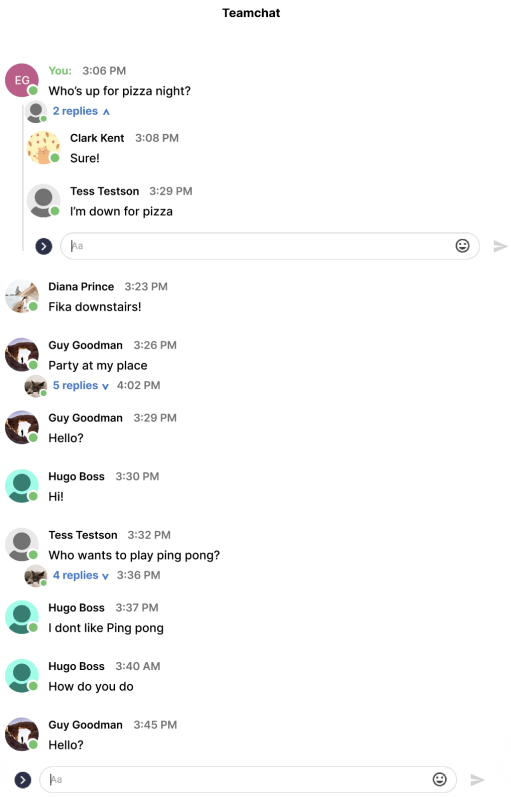

Figure 6.10: Task 4 in version 2.

6.3 Results

The questionnaire answers from each participant were calculated into a final SUS score as described in section 2.4.3. A visualization of the data as well as mean SUS score, standard deviation (SD) and the 95% confidence interval (95% CI) for each group is presented in figure 6.11 and table 6.1. The complete data set can be found in table 6.2.

Table 6.1 shows a relatively large difference of 11 points in mean SUS score between groups 1 and 2, although looking at the 95% confidence intervals it is not possible to say that there is a statistically significant difference between the two groups.

The null hypothesis that there is no difference between the mean SUS score given by groups 1 and 2 (n1 = n2 = 11) was also tested using a one-tailed version of the Mann-Whitney U-test described in section 2.2.3. The test resulted in a U-value of 32 with a p-value of .03288. The p-value is less than the level of significance (alpha = 0.05), meaning the null hypothesis can be rejected and that the mean SUS score given by group 1 is likely smaller than the mean SUS score given by group 2.

Groups 2 and 3 gave an almost identical mean SUS score and the difference is not statistically significant.

Table 6.1: Mean SUS score, standard deviation and 95% confidence interval for each group.

	Group 1	Group 2	Group 3
Mean	74.77	85.91	86.37
SD	13.67	10.18	14.16
95% CI	[66.69, 82.85]	[79.89, 91.93]	[78.00, 94.74]

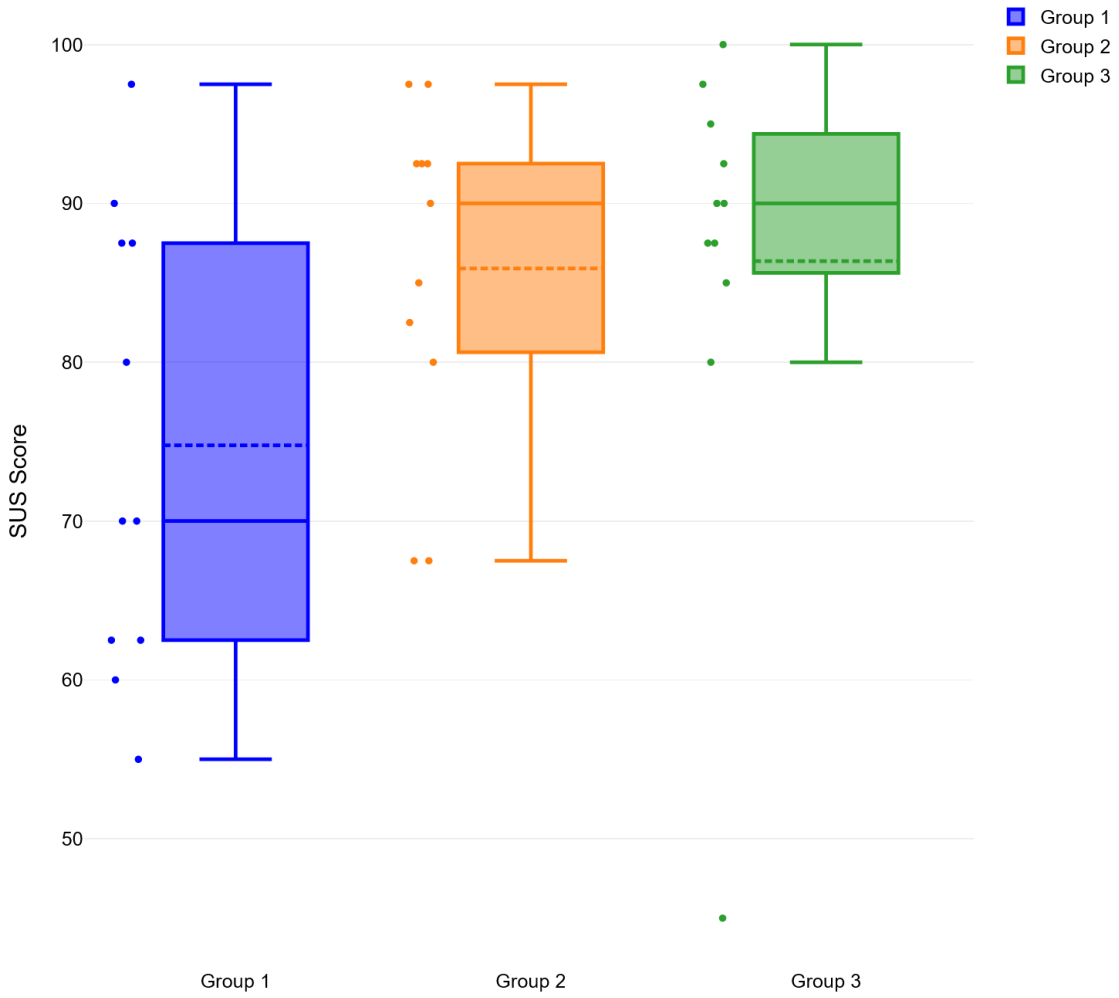

Figure 6.11: Box plots for the SUS scores given by each group. The boxes consist of minimum, first quartile, median, third quartile, and maximum. The dotted line is mean score.

Table 6.2: The SUS score from each participant in groups 1, 2 and 3.

Group 1	Group 2	Group 3
70	97.5	80
55	92.5	92.5
87.5	90	87.5
97.5	67.5	87.5
70	67.5	100
90	82.5	97.5
62.5	80	85
80	92.5	90
60	92.5	45
62.5	97.5	90
87.5	85	95

Chapter 7
Discussion

In this chapter a discussion about the results gathered throughout the design process and from the evaluation of the final prototype is held in relation to the three research questions. The results are interpreted also in relation to the previous work cited in the literature review and methodology. Finally, questions and areas for future research are discussed.

7.1 The Design Process

The results from the evaluation of the final prototype indicate that the application of user-centered design was a successful method for finding and developing an appropriate instant messaging feature for the Telavox application. While it is difficult to pinpoint exactly what parts of the design process contributed to the successful improvement of the application's usability score, there is a discussion to be had about the different strengths and weaknesses in the methodology. What limitations were present during the project, and what could have been done differently given more time and resources?

7.1.1 Initial Product Analysis

To understand the product segment of established collaboration platforms, research about what makes similar platforms such as Microsoft Teams, Slack and Discord popular was necessary as they have competitive products with millions of daily users. Features in these platforms that are missing in the Telavox application were automatically seen as a candidate for future prototypes.

During the project's first interviews, some users mentioned features from other platforms that had been previously overlooked in the initial stages of the design process. A more formal competitive analysis to compare all different applications might have increased knowledge of these platforms, which could then have been applied in the design process to find other more impactful features.

7.1.2 Interviews

All interviews held during the design process were of the semi-structured format. This came with the benefit of allowing flexibility in what areas to focus the interview on. An interviewee with strong feelings regarding a certain topic could be asked follow-up questions, steering the conversation further into that topic and generating a good amount of rich qualitative data. Although the format proved to be a great source of information, there is a risk that the type of data collected could be skewed. Some interviewees were more talkative and social than others, which resulted in high variability in the amount of data collected during each interview. This could have had an effect on the type of thoughts and comments collected during the interviews, as opinions on an application used for communication could very well differ between introverted and extroverted interviewees.

Since interviewees at Telavox were asked to share experiences of the application's potential short-comings as well as features they may prefer in competing platforms, it was important that they felt comfortable during the interview. With this in mind, the goal of the project was stated for the interviewee before each interview session. This created a constructive environment where ideas could flow freely, instead of one where interviewees felt they were criticizing their own product or back-talking colleagues. To further avoid participants feeling uncomfortable, no recording was made of the interview sessions and they were instead transcribed. This resulted in solely relying on notes for the data that emerged during the interviews, which meant it was not possible to get a detailed record of the interview. Another downside of trying to provide a comfortable environment is that it is difficult to provide the exact same conditions for all participants. This could have further skewed the qualitative data gathered.

While there were clearly some negative sides of conducting semi-structured interviews, the benefits outweighed the disadvantages in this case. Using an alternative format like structured interviewing could possibly have resulted in less variability in the quality and quantity of data, but would have generated less qualitative data and would instead have been skewed by what questions were prepared beforehand.

7.1.3 Interview and Test Participants

The group from Telavox that participated in the test and interviews were, as mentioned in section 6.2.1, mostly people that were available at the time. Throughout the project it was difficult to estimate when each prototype would be ready, which, together with the fact that the schedule of the more experienced users was quite busy, meant it was challenging to recruit suitable test participants. Although having some experienced users with more than three years of experience using the platform, the majority of the sample that tested the prototypes held more junior roles at Telavox, or had recently been employed. This might have had a negative impact on the result of the tests. The risk is that the less experienced users have not yet discovered certain usability flaws in the application, or seen what impact it has on their productivity.

On the other hand, having less experienced users in the interviews and tests might have had an overall positive effect on the project. These users have not participated in the development of the Telavox application, neither have they had time to create the same user patterns as some of the more experienced users have. This leads to them potentially having a more un-

biased opinion and perhaps a more objective standpoint in how the application is perceived by users. The reason being that they are not leaving feedback on a product they themselves have built or created and therefore, contrary to the experienced users that might have been involved in the development process, might be more likely to leave honest feedback.

In addition to level of experience, what role at Telavox the participant had might also have had an impact on the outcome of the study. Participants in the study working as engineers, designers, HR and sales have different responsibilities within Telavox and therefore their opinion on what is most important might differ. With this in mind, it would potentially have benefited the end users, Telavox's customers, if more employees working in sales or customer support would have been included in the design process. These employees could possibly have a better understanding of the market and its competitors, as well as what issues and requests the customers have with the current product.

7.1.4 SUS Evaluation

The results of the SUS evaluation showed that there was a statistical difference in the SUS score given to the two versions of the Telavox application. This section will explore and highlight some potential factors that might have had an impact on this evaluation.

The first potential reason for the difference in result is the choice to develop the prototypes in Figma. Although Figma offers a wide variety of tools and functionality to develop advanced prototypes, it is not comparable to a fully functioning application. In this case, prototyping in Figma relied on test participants following an exact path in the prototype when completing a task. With this said, implementing all functionality of the real application in a Figma prototype is unrealistic and is arguably a misuse of the tool itself. As the SUS evaluates the usability of the entirety of a system, limiting the user in what way they can interact with in an application might have an impact on how users experience the application. It could therefore be argued that doing a real implementation instead of a Figma prototype would allow the users to explore the application more freely, giving them a more true experience and thus yielding a more reliable SUS score.

Another perspective worth mentioning in the context of using Figma as a prototyping tool is that performance of the new feature was never measured. The test was designed entirely around measuring the preference of the users. The main reason to this was that the test participants had varying experience with how Figma prototypes works. Some had used the tool in a professional context and some had never heard of it. Measuring performance such as time to complete a task or number of tasks completed would therefore put some participants at a disadvantage which could potentially give a misleading result.

Another factor to look at during testing of the final prototype was that the scenarios and tasks created for the SUS evaluation somewhat disfavored the original version of the application, and could therefore have negatively affected the SUS score for that version. Although this was necessary for demonstrating situations where it was possible to use the reply feature, adding a scenario where the new feature instead is disfavored could potentially have altered the outcome of the SUS evaluation.

The last factor recognized to have impacted the result from the SUS evaluation is that all the test users might not have understood the questions in the SUS questionnaire, or what it was meant to investigate. Time was spent trying to make the questionnaire as clear as possible, but the language barrier for non-native English speakers was a hurdle in a few cases.

7.2 Answers to Research Questions

RQ1: What instant messaging features are essential for good usability in a collaboration platform?

The prototype with the new reply feature scored significantly higher in the SUS evaluation compared to the original version without the feature. The significant improvement in SUS score speaks for that the new reply feature could be considered an essential instant messaging feature. While it is hard to directly generalize the result, a question worth considering is if a similar feature would have the same effect on other platforms.

Some other possible candidates that could be named "essential" features emerged from the qualitative data collected during the initial exploratory stages of the design process. Functionality for organizing messages was a theme with wide representation in the data, which helped steer the project into designing the reply prototype. Moreover, functionality for organizing chats was also a strong candidate, with many users mentioning a large amount of different chats necessitated some way of personally organizing these for each user's own needs. Another heavily requested feature was the ability to search for information in chats, which was mentioned as a desirable option for tackling issues with poor chat structure and findability. A common characteristic amongst all these features is that they allow the users more freedom in how they can organize and find information in the chat.

While the results suggest that some of the above-mentioned features could be considered essential, it should be noted that this depends on the needs of the target user group, as well as what goals are set for the product. If the goal is to use the platform for internal use or as a nice-to-have application that comes with other products, a collaboration platform with the most basic features to ensure communication between employees might be sufficient. It is important to remember that Telavox employees have been using the platform for many years even without these features. If the goal is instead to provide a competitive product that should cater to the needs of many different kinds of users, one can make the case that it requires these features in order to compete with already popular collaboration platforms.

To get a more definitive answer on what features should be considered essential for good usability, a larger study on multiple platforms with many more groups of users and larger sample sizes would be desirable.

RQ2: How much can an instant messaging feature improve the perceived usability of a collaboration platform?

The results of the SUS evaluation shows that the added reply feature significantly improved the perceived usability of the Telavox application. While it is difficult to quantify the improvement due to the high degree of uncertainty around the aggregate score given by group 1, comparing the mean score of groups 1 and 2 indicates that the improvement could be as large as 11 points. This would mean that the system usability was improved from a SUS score of 75, to a more superior score of 86. It would have been preferable to also compare group 3 to a fourth group consisting of Telavox employees testing the application without a reply feature. Concerns arose, however, that this group would be too focused on the limitations of the Figma prototype as they were used to the real application.

In an earlier study, a search feature added to the same application yielded similar results with a SUS score of 93 based on a sample of eight Telavox employees [23]. While the search feature is not very similar to a reply feature in its functionality, both features do try to counter the issue of findability in the application's chat system. Perhaps it can then be argued that this result further validates the result seen in this study, and gives more weight to the argument that an individual feature can have a large impact on the overall perceived usability of a collaboration platform.

While the results suggest that the new reply feature improved the usability of the collaboration platform, there are some limitations to the methodology. To begin with, only one method to measure usability was utilized for the study: the SUS evaluation. Additional methods and evaluations could give more depth and insight on how this particular change impacted the usability of the application. As mentioned earlier in section 7.1.4, it would be preferable to also measure the performance of participants completing tasks in the prototype. Moreover, the prototype versions used in the tests are not certain to have provided an accurate experience of the real Telavox application to the participants, which could have further affected the result.

RQ3: How are new features received by experienced versus inexperienced users?

According to previous research, prior experience in a system can influence the usability score given by a user by as much as 6-15 percent [16]. The expected outcome of the SUS evaluation in this study was that there would be a similar difference in the perceived usability between the two groups of experienced and inexperienced users. The hypothesis was that familiarity with the Telavox application would let the experienced users quickly adapt to the new feature and instead focus on how it could counter the workflow issues shared by their colleagues during the design process. In contrast, the inexperienced users were expected to be more focused on familiarizing themselves with the application and its core features.

The results show that there was no significant difference in the SUS scores given by experienced and inexperienced users. This suggests that the new prototype reply feature may be equally accessible and useful to both groups of users, regardless of their level of experience with the platform, indicating a good level of intuitiveness and ease-of-use. Interestingly, the design process was almost exclusively centered around feedback from the experienced users. One possibility is that the needs and goals of the two groups are more similar than expected. Perhaps experience in a similar platform could be regarded as equal to experience in the Telavox application, though it is not possible to confirm this since users without experience in any collaboration platform were excluded from the test.

Overall, the results indicate that the new prototype reply feature is a valuable addition to the platform that can be effectively utilized by both groups of users. It is worth considering, however, whether further qualitative user testing with a larger sample size could provide more insight into other differences between the needs and goals of these two groups.

Furthermore, the sample size may have been too small to accurately represent differences in the needs and goals between experienced and inexperienced users. A larger sample size may have resulted in more statistically significant differences between the two groups. Additionally, it is possible that the results may have been different if for example the experienced group had been divided into an intermediate and advanced group.

7.3 Future Work

Based on the results and findings in this thesis, there are several paths to pursue in future research. The most natural direction that this project has laid the groundwork for would be the investigation of a real implementation of the reply feature in the Telavox platform. This could itself open up the opportunity to research other metrics such as the general user experience of the platform, which would generate a more nuanced view on how a feature can improve a platform. Other user groups like intermediate users and those without *any* experience using collaboration platforms could also provide further insight into how different backgrounds influence the perceived usability of a platform.

Another direction that could yield interesting results based on this study is implementing or prototyping one of the other features that the initial research suggested, or other unrelated features. Those features could then also be compared with the results of this study to provide more insight into what impact different kinds of features can have on the platform.

Finally, an interesting issue that could potentially be researched further is how to balance the requirement of limiting space given to unnecessary messages with increasing the space given to important messages. For example: is it possible to distinguish between such messages on a per-user basis?

Chapter 8
Conclusion

The goal of this thesis was to investigate the impact of instant messaging features on the usability of collaboration platforms. This was achieved through an iterative design process in which usability testing and user feedback helped shape new prototype features for the Telavox application. Established methods connected theoretically supported concepts to users' goals and needs with a collaboration platform.

The findings suggest that instant messaging features that address issues of organization, structure, and findability can be essential for good usability in collaboration platforms. Adding a reply feature to the Telavox application gave a significant improvement to the perceived usability of the application, and it is possible that a similar feature could yield similar results in other collaboration platforms. However, it should be noted that the specific features that are considered essential may vary depending on the needs and goals of the target user group.

How impactful a feature can be to the overall usability of the platform was demonstrated by the significant increase in the SUS score correlated with the introduction of the new reply feature. The result of the SUS evaluation also concluded that there was no significant difference between how experienced users and inexperienced users perceived the new feature, perhaps indicating that experience in similar collaboration platforms could be considered equal and that lack of experience does not necessarily affect how a user will perceive the usability of a platform. The successful improvement of the platform could also be said to further attest to the strength of the user-centered design process as a method.

While the thesis yielded a result with statistical significance, the exact number of the increase in SUS score was difficult to ascertain due to the small sample size and high variance in the samples. There were also circumstances related to the methodology that might have impacted how the users perceived the prototypes used in the design process, which may thereby have influenced the final result.

Future research could build on the findings in this thesis by investigating how a real implementation would fare in a similar study with a larger sample size and perhaps other user groups. Another interesting angle would be to implement and explore the impact of other features and compare them with the results in this study.

Bibliography

[1] A. Bangor, P. T. Kortum, and J. T. Miller. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, 24(6):574–594, 2008.

[2] J. Brooke. Sus: A "quick and dirty" usability scale. *Usability evaluation in industry*, pages 189–194, 1996.

[3] K. McCartan C. Robson. *Real World Research*. Wiley, 2nd edition, 2016.

[4] E. Charters. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal*, 12(2), 2003.

[5] D. Curry. Microsoft teams revenue and usage statistics, 2022. https://www.businessofapps.com/data/microsoft-teams-statistics/ (accessed Dec. 11, 2022).

[6] J. S. Dumas and J. Redish. *A practical guide to usability testing*. Intellect books, 1999.

[7] G. Nyktarakis. Technology Enabling Collaboration of Agile Development Teams in Hybrid Working: The case of Microsoft Teams, 2022. Student Paper, Faculty of Technology, Linnaeus University, Sweden.

[8] J. D. Gould and C. Lewis. Designing for usability: key principles and what designers think. *Communications of the ACM*, 28(3):300–311, 1985.

[9] J. Gubrium, J. Holstein, A. Marvasti, and K. McKinney. *The SAGE handbook of interview research: The complexity of the craft*. Sage Publications, 2012.

[10] ISO. 9241-11:2018 ergonomics of human-system interaction — part 11: Usability: Definitions and concepts, 2018. https://www.iso.org/standard/63500.html (accessed Jan. 12, 2023).

[11] H. Hochheiser J. Lazar, J. Feng. *Research Methods in Human Computer Interaction*. Morgan Kaufmann, 2nd edition, 2017.

[12] Y. Rogers J. Preece, H. Sharp. *Interaction Design: Beyond Human-Computer Interaction*. Wiley, 4th edition, 2015.

[13] Y. Rogers J. Preece, H. Sharp. *Interaction Design: Beyond Human-Computer Interaction*. Wiley, 5th edition, 2019.

[14] J. R. Lewis. The system usability scale: Past, present, and future. *International Journal of Human–Computer Interaction*, 34(7):577–590, 2018.

[15] H. Mann and D. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60, 1947.

[16] S. McLellan, A. Muddimer, and S. C. Peres. The effect of experience on system usability scale ratings. *Journal of usability studies*, 7(2):56–67, 2012.

[17] United Nations. Transforming our world: the 2030 agenda for sustainable development. https://sdgs.un.org/2030agenda (accessed Dec. 14, 2022).

[18] C. Neiger. Microsoft teams has 13 million daily users, 2019. https://finance.yahoo.com/news/microsoft-teams-13-million-daily-200000300.html (accessed Dec. 11, 2022).

[19] J. Nielsen. How to conduct a heuristic evaluation. *Nielsen Norman Group*, 1(1):8, 1995.

[20] J. Nielsen. Ten usability heuristics, 2005. https://www.nngroup.com/articles/ten-usability-heuristics/ (accessed Jan. 5, 2023).

[21] M. Rettig. Prototyping for tiny fingers. *Commun. ACM*, 37(4):21–27, 1994.

[22] V. Stray and N. Moe. Understanding coordination in global software engineering: A mixed-methods study on the use of meetings and slack. *Journal of Systems and Software*, 170:110717, 2020.

[23] T. Glas and S. Christiansson. Improving User Experience in Chat Tool by Implementing an Intuitive Search Feature, 2022. Student Paper, Faculty of Engineering, Lund University, Sweden.

[24] T. Warren. Microsoft teams usage jumps to 145 million daily active users, 2021. https://www.theverge.com/2021/4/27/22406472/microsoft-teams-145-million-daily-active-users-stats (accessed Dec. 11, 2022).

[25] C. Wilson. *Brainstorming and beyond: a user-centered design method*. Newnes, 2013.

Appendices

Appendix A
SUS Questionnaire

System Usability Scale

© Digital Equipment Corporation, 1986.

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Figure A.1: The SUS questionnaire from Brooke [2] used in the SUS evaluation.

Appendix B
Interview Schedules

B.1 Interview Schedule (Initial Product Analysis)

Introfrågor:

1. Fråga om ålder och roll på Telavox. Hur länge har du använt meddelandesidan i Telavox-appen?
2. Vilka av följande chatt-applikationer har du använt i ett arbete tidigare? Microsoft Teams, Discord, Telavox, Slack.
3. Vilken av dessa tyckte du bäst om att använda, bortsett från Telavox-appen? Varför?

Huvudfrågor:

4. I relation till de föregående apparna, har de något som saknas i Telavox-appen? Varför är det viktigt? Något annat som borde finnas?
5. Finns det något (funktion eller gränssnitt) i Telavox-appen du inte gillar? Varför?
6. Följd till förra frågan: Något som behöver förbättras? Hur då? Exempel på bättre implementationer?

Om användning av olika chattar för olika ämnen och information:

7. Hur använder ditt team meddelandesidan i ert arbete? Utveckla med alternativ: Flera chattar, privata meddelanden, olika chattar för olika sprints och issues?
8. Finns det utrymme att förbättra detta på något sätt? Vad kan förenkla arbetet för teamet?

Om rum vs chattar:

9. (Om rum/chatter inte nämnts) Har du använt funktionen för att skapa ett rum?
10. (Om svararen inte vet vad det är: Visa hur man skapar ett rum. Exempel på publika rum.) Vad tycker du om möjligheten att skapa rum och vanliga chatter? Försök leda in på hur lätt svararen har att förstå funktionen. Finner hen det användbart?

Avslutande:

11. Hur hade du rangordnat de förslag vi diskuterat? Viktigast till minst viktigt?

B.2 Post-test Interview Schedule (Iteration 2)

Upon completion of the tasks the following questions were asked.

1. Do you have any questions about the prototype?
2. Did anything feel unclear or unintuitive?
3. Is there anything you want us to explain further?
4. Was there anything you liked and/or disliked about this prototype?
5. Would you want to change anything about this prototype?
6. Can you imagine any scenarios where this feature would have been helpful?
7. Can you see any downsides in having this feature?
8. Would you use this feature if implemented in the chat?

Possible follow up questions:

1. Why did you like/dislike that particular part of the prototype?
2. Would the upsides of implementing this feature make up for its possible downsides?
3. How often would you use it on a daily basis?
4. In what way could this increase your productivity?

Appendix C
Final Prototype

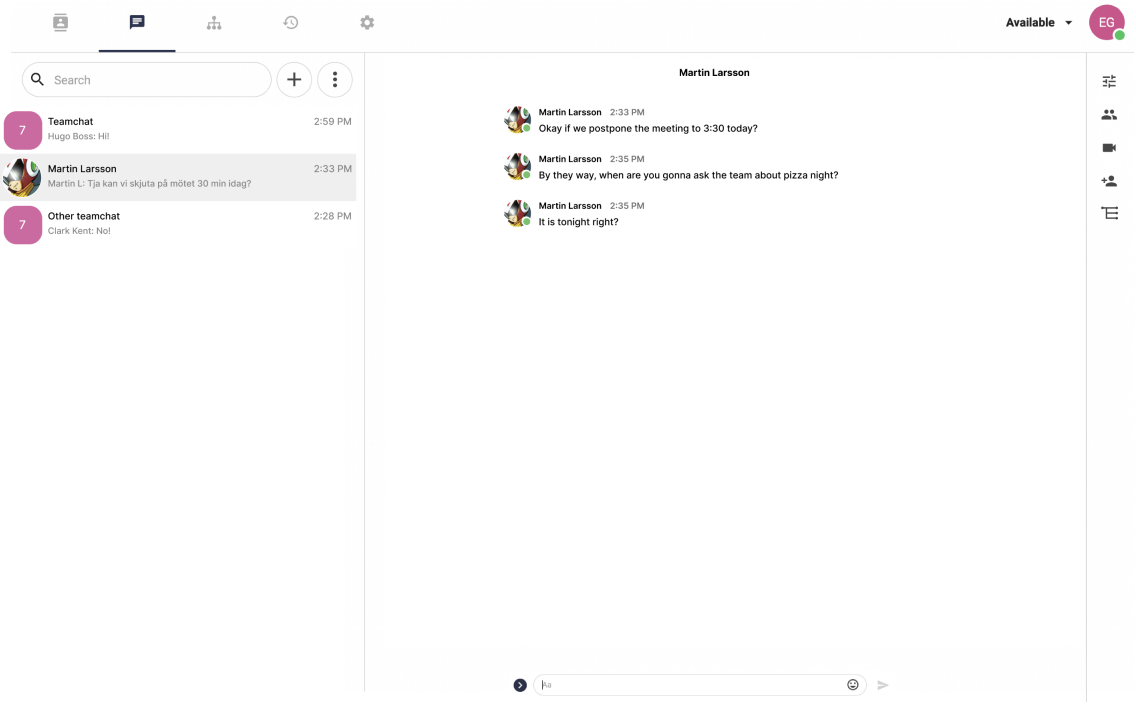

Figure C.1: Start of task 1 in the final prototype.

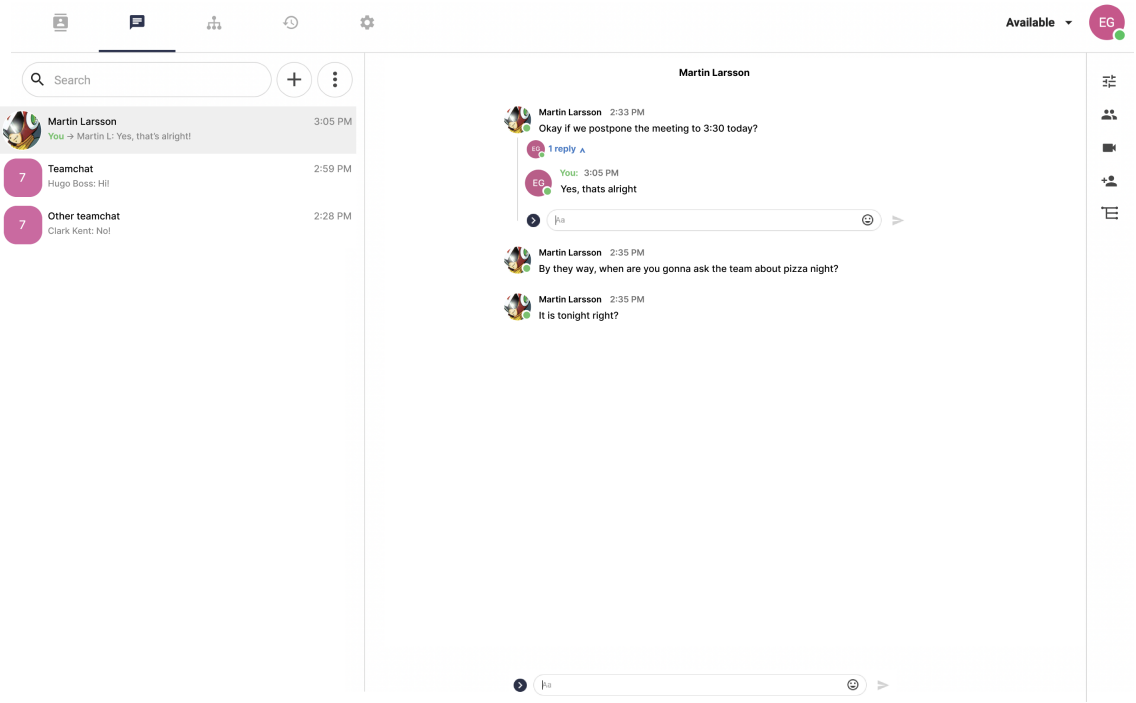

Figure C.2: A message reply in task 1.

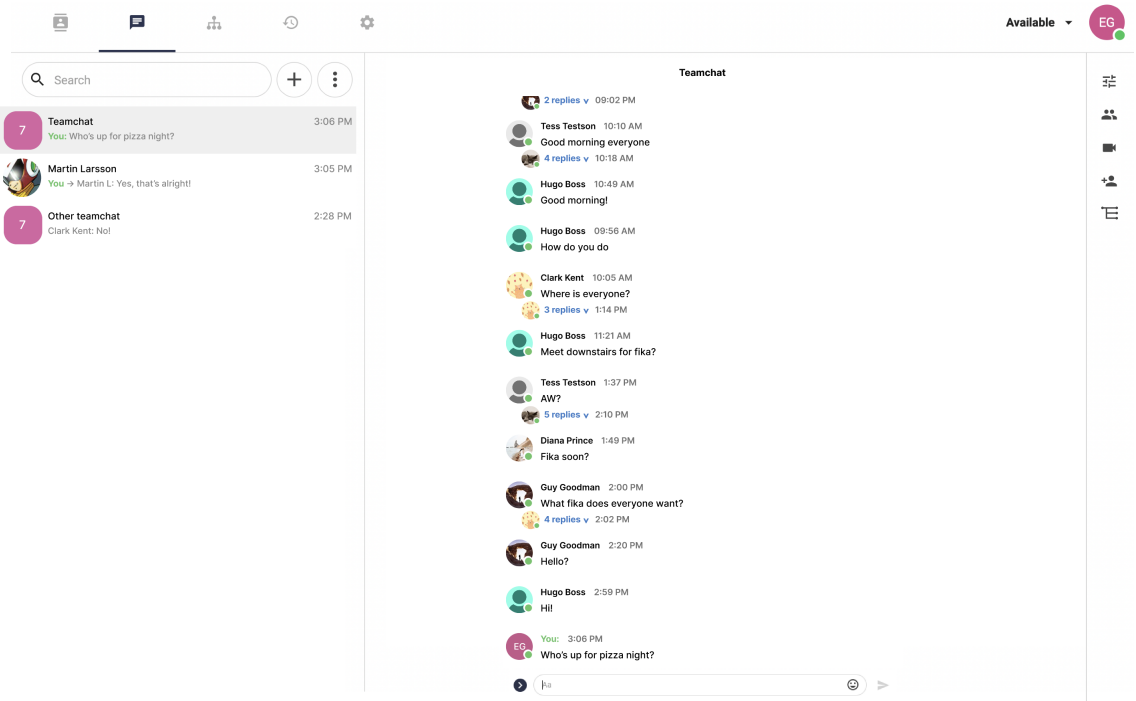

Figure C.3: A normal chat message in task 2.

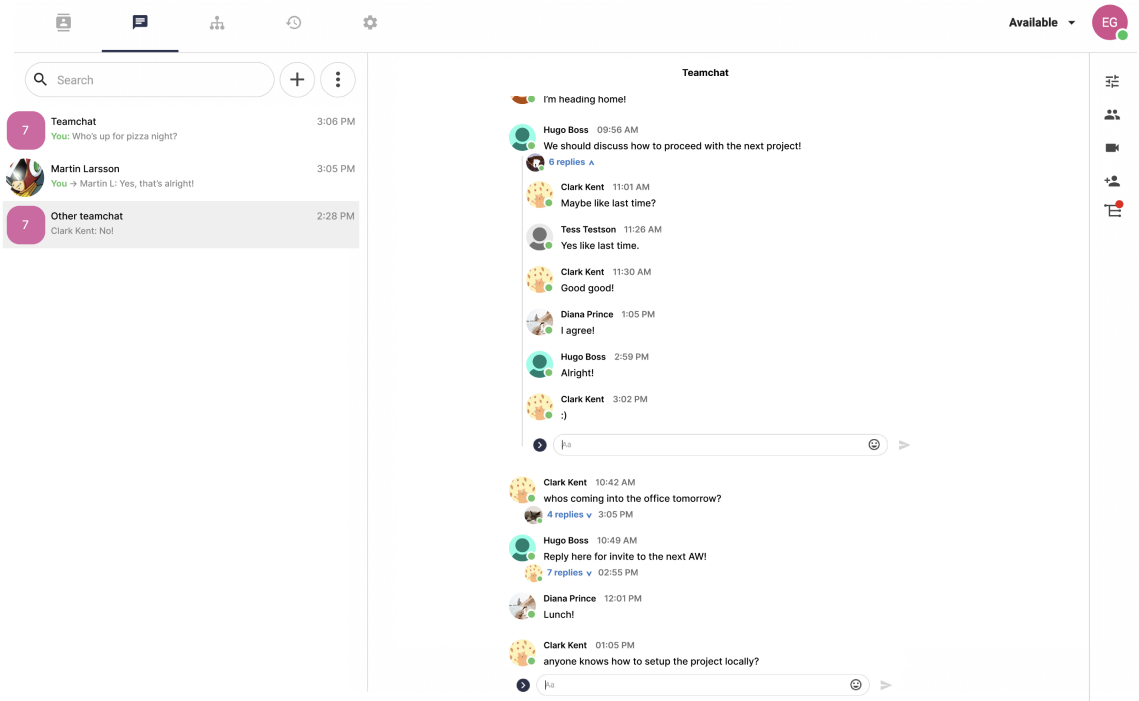

Figure C.4: Task 3.

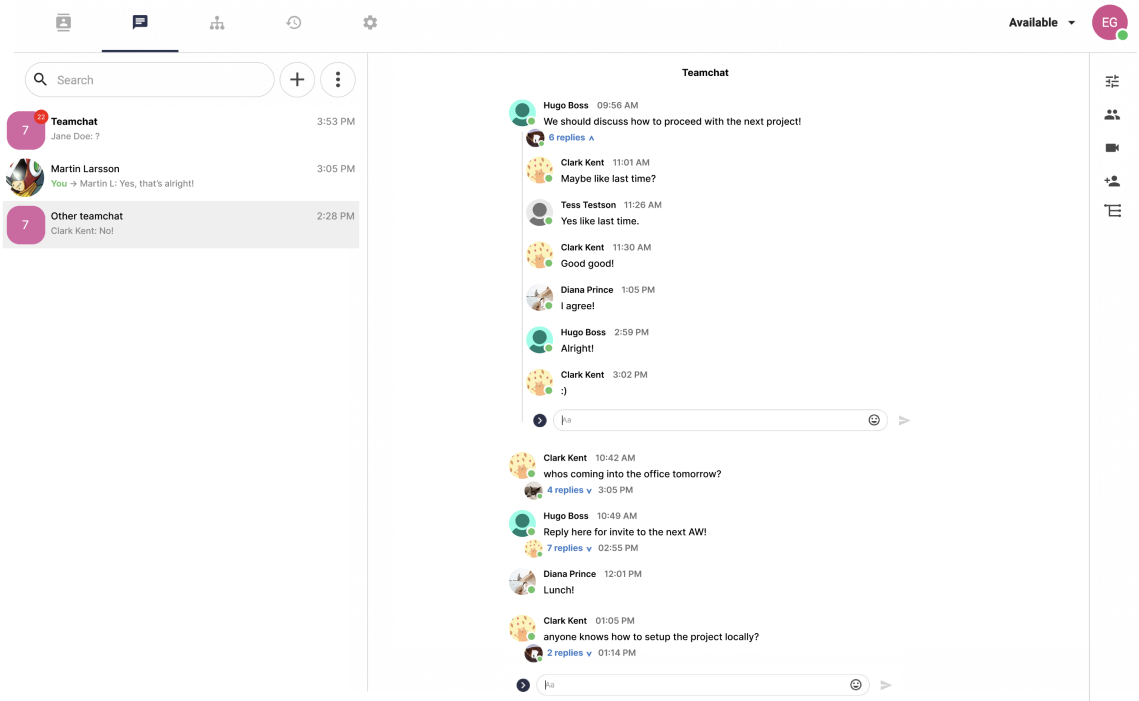

Figure C.5: Start of task 4.

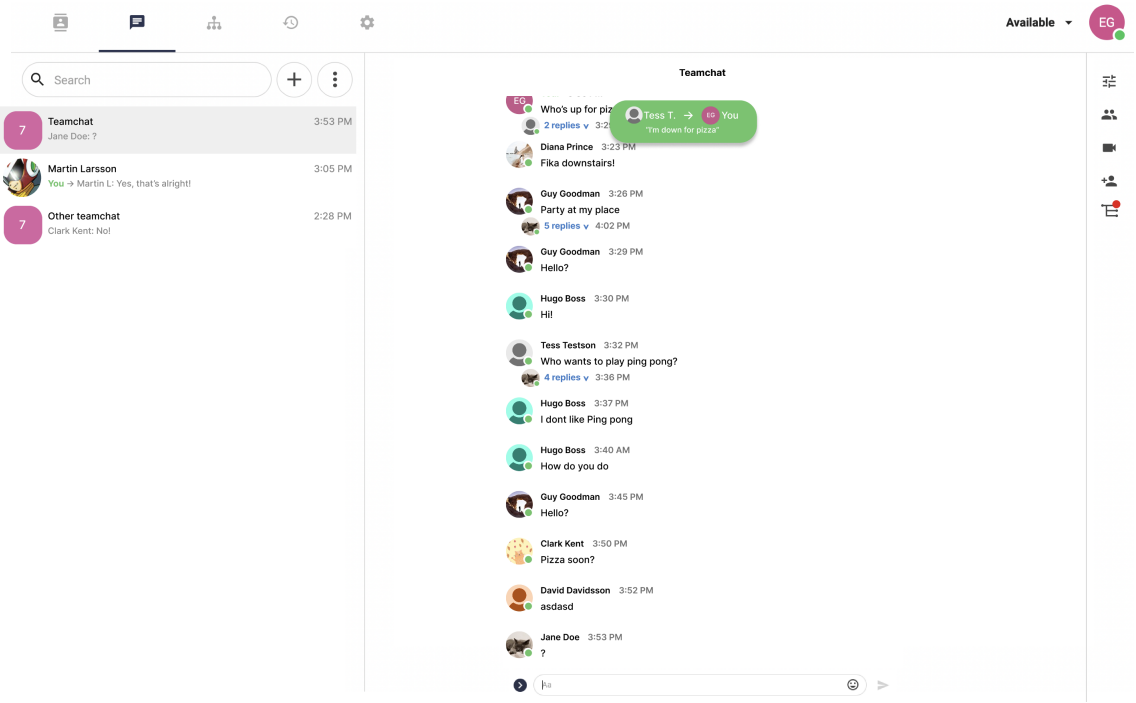

Figure C.6: Pop-up notification in task 4.

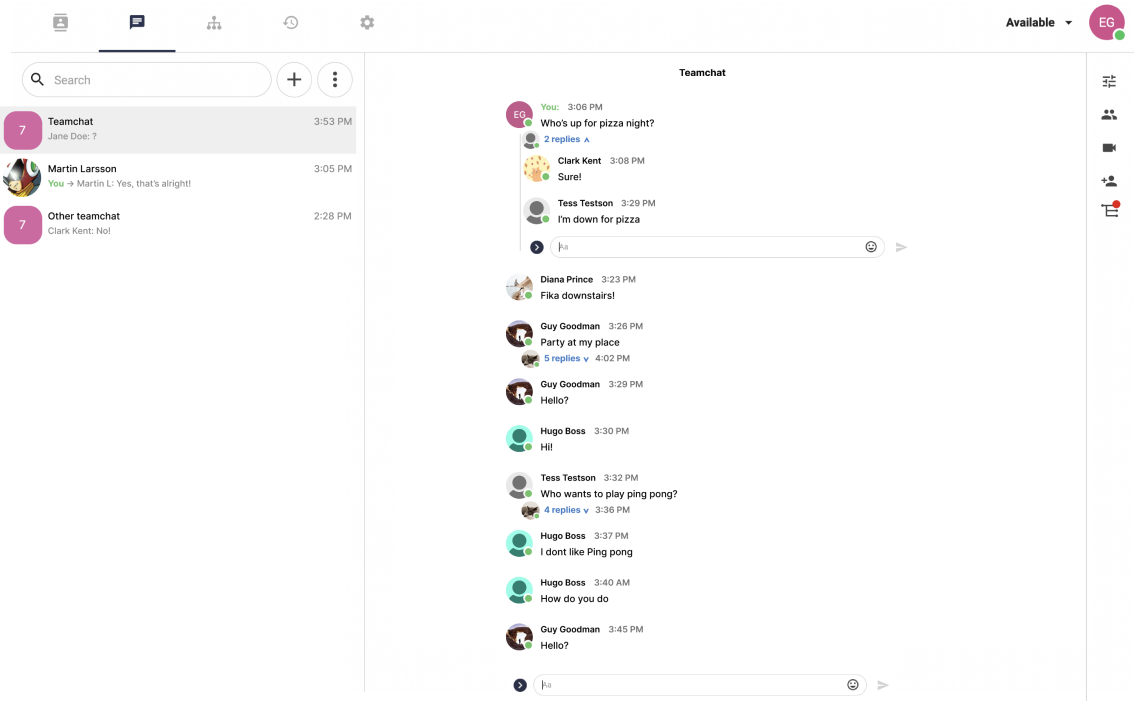

Figure C.7: End of task 4.