University of Waterloo
Faculty of Engineering

Usability Tests for the New Interests Feature

Nextdoor
875 Stevenson Street #700
San Francisco, CA 94103, USA

Prepared By
Ashley Changyi Hu

May 14, 2018

Abstract

With the emergence of a new Interests feature, Nextdoor hopes to conduct a usability study to help them better understand how their users will respond to this feature. A series of usability tests were executed to determine the presentation of Interests that is most commitment-free, light-weight, and clear to users. The variants of the study include “Add to Feed”, “Subscribe”, “Follow”, and “Join”. A supplementary A/B test was additionally introduced to inform the results from the initial 2-part usability tests. From synthesizing the qualitative and quantitative results, “Join” and “Follow” seemed to be the best performing call-to-action in both cases. An executive decision was made to move forward with “Join” as it more closely aligns with how Nextdoor wants subscribers to interact with the Interests feature. Specifics related to the testing process and data analysis will be discussed within the contents of the report.

1

Introduction

1.1 Understanding Usability Testing

In most software companies, usability testing or user testing has become an important part of product workflows. Often times, products are designed with assumptions about user behavior in mind. These assumptions, although usually backed up by valid reasoning, may be right or wrong. The objective of user testing is to create a set of actionable tools that teams can use to validate these decisions, which are being made on a daily basis [1]. Combining both qualitative and quantitative research methods, designers use findings from these tests to build product experiences that reflect the mental model and needs of the people using it [2]. In short, it helps a company, like Nextdoor, to better understand their users.

1.2 Situation of Concern

Nextdoor recently released a new feature called “Interests”, which are essentially groups that users can join. Some examples of Interests would be biking, hiking & trails, and gardening. Once joining an Interest, neighbors part of the group would then receive notifications when there are activities, events, or posts related to this Interest. This is to say, each neighborhood has their own Interest groups, and neighbors are restricted to Interests only within their own neighborhood. The purpose of the newly introduced feature is to build a sense of community amongst neighbors through encouraging meetups and triggering conversation.

When promoting this new feature, the designers struggled to determine the best call-to-action to gain user traction. Some of the options that were up for debate include: add to feed, subscribe, follow and join. Since the feature was predicted to significantly grow membership across the entire product, the decision being made here was valued at high

priority. Thus, this called for a user test to validate the various call-to-action options, along with a designer to take responsibility for the synthesis of results.

2

Usability Test Design

2.1 Goals and Objectives

There are a few main goals that the membership usability test aims to achieve. The first is to identify the call-to-action (CTA) that makes the Interests feature feel the most lightweight and commitment-free. The second is to determine the CTA that most accurately aligns with our users' mental model on what to expect upon joining a group. Finally, the goal is to also find the membership CTA that introduces the most subscribers to the Interests platform. This in turn should boost user retention for Nextdoor's entire product.

These three goals will constantly be referred to throughout the design and synthesis of the test as to not detract from the focus of the study. Breaking the objectives down into a clear set of questions, the following should be answered through conducting the test:

- Which verb is most likely to get the most clicks: add to feed, subscribe, follow, or join?
- Do any of these help users understand what to expect and what the value of Interests is better than the others?
- Do any of these cause confusion surrounding the feature itself?

2.2 Requirements and Constraints

Although requirements were not explicitly stated by the product manager overlooking the tests, a few requirements could be concluded from simply meeting with other stakeholders on the project:

- The usability test should meet the three goals stated previously. If no conclusion can be made, a different test will need to be conducted until there is a firm winning solution.

- The structure of the usability tests should eliminate as many biases as possible in order to ensure the validity of the study.
- Any screens used for the tests should reflect the true nature of the product as it appears to the rest of the users on the platform. Any visuals should also abide by the usability and brand guidelines established by the design team at Nextdoor (if relevant).

In addition to the requirements above, there were a few constraints that limited the scope of the user tests conducted:

- The study should not exceed the budget allocated to it. As the project is considered to be one of high importance, this factor hopefully should not play a huge role.
- The test itself should not reveal any new features on Nextdoor's product that have yet be released to the public, as specified in each employee's Non-Disclosure Agreement.
- The tests should not exceed the 4-week deadline set for the project. It should actually only take an estimated time of 2 weeks to allow sufficient time for analysis of results.

3

Designing the Usability Test

3.1 Analyzing User Insight Methods

Depending on the needs and restrictions of each project, a user test can take on many forms. To name a few common testing methods, the following are ordered from generative to evaluative types of tests: ethnography, participatory design, concept testing, surveys, usability study, A/B testing, analytics, and net promoter score (NPS). A key consideration to choosing the right method is to understand how each is different from the rest, as well as the type of analysis that can be abstracted from the results. Table 1 below highlights some of these differences. [3]

Table 1: Summary of Common Usability Testing Methods [3]

Method	What is it?	Great for…
Ethnography	In context, in-depth, observations and user interviews	Revealing new opportunities and latent needs, building empathy and "informed intuition"
Participatory Design	Structured activities to engage users in ideation and product design	Sparking unexpected ideas, gaining insight into user needs and desires
Concept Testing	Eliciting qualitative preferences and feedback about a range of product ideas, flows or designs	Quickly prioritizing features, understanding the relative desirability of several value propositions
Surveys	Simple way to digitally gather qualitative and quantitative information	Assessing preferences, attitudes, characteristics and opinions on a given topic

Usability Study	Observing user's attempts to accomplish various tasks in the product or prototype	Evaluating product efficacy, identifying areas of improvement
A/B Testing	Quantitative recording of user behaviour across product variants	Optimizing product effectiveness
Analytics	Quantitative analysis of user behaviour and patterns in the live product	Uncovering gaps between actual vs. desired behaviours, identifying opportunities for improvement
Net Promoter Score	A method for measuring brand or product loyalty	Correlating new features, behaviours or demographics with user satisfaction, gathering qualitative feedback

The methods described above are not mutually exclusive and can be used in conjunction with one another [3]. Each method also reflects its own pros and cons; thus by understanding the differences, the most suitable method can be chosen to accommodate for the initially defined goals and estimated resources available.

3.2 Determining the Best Method

Referring back to the goals of the study, the first step is to determine whether the usability test should take a qualitative or quantitative approach. Generally, qualitative tests answer the questions "why" and "how". By contrast, quantitative tests focus on the "what" aspect of a study [2]. The below graphic was created by Nextdoor's design team to depict the categorization of various research methods in terms of how qualitative or quantitative it is. It also takes into consideration of the distinction between behavioral and attitudinal studies, which simply differentiates between what users actually do from what they say or think they do. [3]

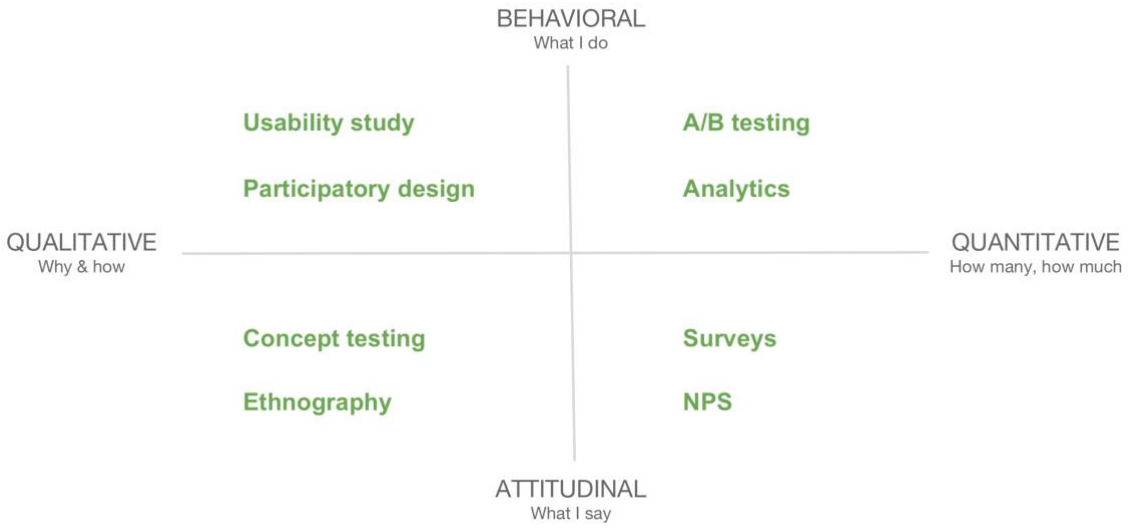

Figure 1: Categorization of Test Methods on a 2-D Spectrum

Ideally, results from a quantitative test should complement that of a qualitative test [1]. Referring back to the goals, it became clear that although the project was not limited in terms of money, it was instead restricted by time. A 4-week period is quite a short timeline for a complete usability test, especially with other ongoing projects running simultaneously. Therefore, this meant that any form of a moderated test would be out of question. Moderated tests call for time to recruit participants, along with additional time to actually set up and later synthesize the study. In addition, the act of moderating itself requires nuance and skill, all of which are resources that Nextdoor currently does not have to complete such a test in the amount of time allocated.

To ensure that the most valuable and actionable results are obtained from the usability tests, the design team still believed in the benefit of incorporating a combination of all four aspects on the 2-D spectrum (qualitative, quantitative, behavioural, attitudinal). The trade-off here is to do so in the simplest and least time-consuming way possible.

Decision-Making Process

Since the results of the project would impact multiple project teams, the decision-making process welcomed the input of others in the form of open discussions. Throughout the course of the project, three informal meetings were called upon, where designers, product

managers, and engineers interested in the project gathered to explore and discuss the "best" methods to conduct the usability test. In a similar manner, these meetings were used to share any progress or findings related to the study as the tests moved forward.

To kick off the initial discussion, everyone in the room was asked to complete a decision matrix similar to the one below.

Table 2: Modified Decision Matrix for Screen Prioritization [4]

Method	*Most* Value from test insights	*Least* Effort to conduct and analyze	*Least* Time consuming	Total
Weight	**2**	**1**	**1**	
Rating: Ethnography	3	1	1	
Score: Ethnography	6	1	1	8
Rating: Participatory Design	2	1	2	
Score: Participatory Design	4	1	2	7
Rating: Concept Testing	2	2	2	
Score: Concept Testing	4	2	2	8
Rating: Surveys	1	3	3	
Score: Surveys	2	3	3	8
Rating: Usability Study	3	2	2	
Score: Usability Study	6	2	2	10
Rating: A/B Testing	2	3	3	
Score: A/B Testing	4	3	3	10
Rating: Analytics	2	2	1	
Score: Analytics	4	2	1	7
Rating: Net Promoter Score	1	1	1	
Score: Net Promoter Score	2	1	1	4

Focusing on the three factors (value, effort, time), each testing approach is evaluated on a scale from one to three, where one is least optimal and three is most optimal. These values are then multiplied by the weight assigned to each quality, resulting in an overall

score. It should be noted that the options are ranked relative to our project goals. While a participatory test would be very beneficial for the purpose of idea exploration, it is less useful in our case since the objective is simply to back up the pre-identified CTA options with metrics [3]. In this sense, higher totals in the matrix correlate to the methods that more closely aligned with our needs and constraints.

For the most part, most people were on the same page in terms of their ratings. Any major discrepancies were resolved with discussions, where the team walked each other through their reasonings until a mutually-accepted value was settled upon. From completing the matrix together, the obvious winners were the usability study and A/B testing methods, leading the remaining solutions by at least two points.

Final Plan

Following the meetings with other Nextdoor members, a conclusion was made that an A/B test in conjunction with a two-part usability test seemed like an ideal solution to achieving the set-out goals within the constraints of the project.

Briefly planning out the structure of the user tests, part one of the study would include unmoderated usability tests, where various groups of users would be given the same set of tasks to accomplish on different user interfaces. From observing how users think and interact with alterations of the platform, a qualitative analysis can be pulled from the results [1]. The goal of part one of the test is to try to understand “how users feel” as they complete the tasks and think out loud as they do so.

The second part of the test aims to relate “how users feel” and “what they say they do” to how they actually behave on the platform. In the study, this takes the form of A/B tests which complement the prior usability tests. Often, there is a gap between the two, as people are not always accurate in how they perceive their natural thoughts and emotions [4]. Here, it is essential to remember that biases should be taken into account when synthesizing the final results.

4

Executing the Plan

4.1 Creating the Prototypes

Prior to writing the test script, prototypes should first be created so that all possible combinations of test variants are included. Variants simply represent the variations of screens seen by the users participating in the test. For example, if user A sees variant 1A, the user will not see variant 1B. The variants are mutually exclusive, where one of them acts as a control. This makes it easier to pinpoint the “best” solution.

For this project, the design team was tasked to try out a new online prototyping platform called Marvel.

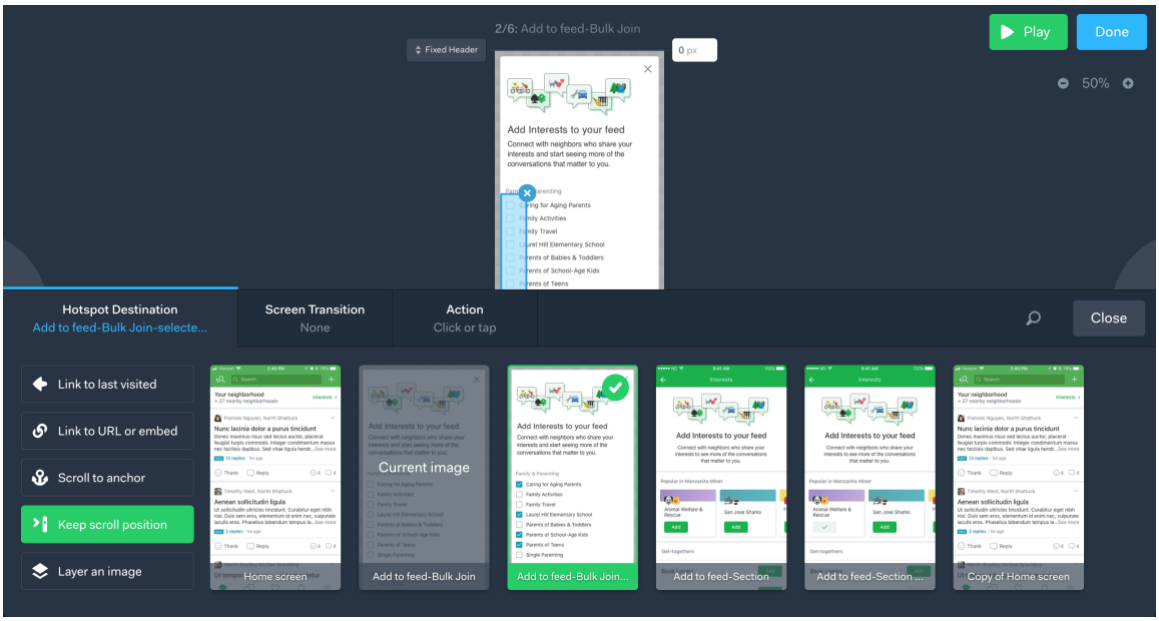

Figure 2: Setting Up Prototypes on Marvel

Designing the prototypes based on the plan, users are shown both of the following two screens on Marvel and are encouraged to talk through their thought process as they complete a set of tasks.

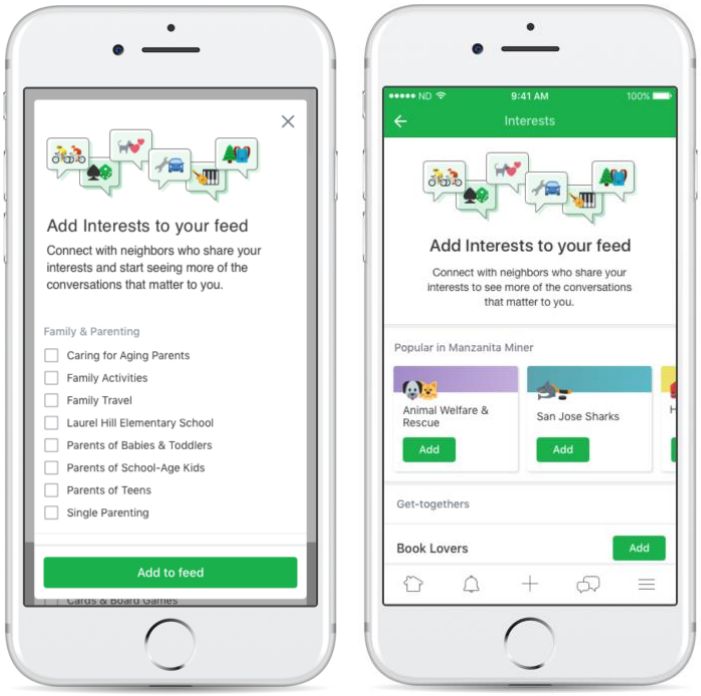

Figure 3: Bulk Join / Screen A (Left), Section Join / Screen B (Right)

In Task A, users are directed through the bulk join flow on the left. In Task B, which takes place after Task A has been completed, they are asked to click through the section join flow. From the users' comments throughout these tests, it can later be analyzed how they "feel" about Task A in relation to Task B. Questions that should be interpreted along the way include "What is their understanding of the new feature" as well as "Which flow appears to be the simplest and most commitment-free". Since the Interests feature can be presented visually in the two (amongst the many other) ways, the purpose of the tests is to support the understanding of these questions, as described previously in the goals.

To further understand how users interact with Interests, another portion of the test investigates the language used in the green CTA buttons. Although all users are exposed to the above two screens, each test group encounters a slightly modified version of it. Some users see "Add to feed" on the first screen (Task A) along with the corresponding

"Add" buttons on the next screen (Task B), whereas other variants may see "Subscribe", "Follow" and "Join" on both screens instead. In other words, all groups experienced variations of the same test flow.

4.2 Writing the Script

When it comes to writing the script, many factors should be considered to ensure that the goals of the test are met. As well, drafts of the script should be reviewed by a variety of people to remove biases that can potentially be avoided. Thinking back to the goals discussed at the beginning of the report, there are three main aspects to be determined:

- Which presentation of Interests is the most light-weight?
- Which presentation of Interests is the most commitment-free?
- Which presentation of Interests is the clearest (or least confusing)?

These questions were continuously referred to while drafting up the script. In addition to simply aligning the script with our planned objectives, it is also important to keep a few considerations in mind:

- *How do we prevent participants from feeling "stuck" while completing the tasks?* Although it is normal to feel lost during a user test, it is not a good sign if they do not understand what is being asked or how to even begin to approach the tasks assigned. This will only result in their frustration, which may change how they perform throughout the remainder of the test.
- *When should we be specific about the task versus when should we allow participants to explore the feature? What are the benefits of each and how does that align with the goals we set?* A comparison can be made between the tasks "Click the on the Interests button" and "Navigate to the Interests section", which does not indicate to the user how the action is completed.

In either case, the above points should be taken into consideration prior to finalizing the test script.

Part 1 of the Test: Task A

The first part of the test aims to provide users with the opportunity to express how they feel and what they understand about the new Interests feature. This takes form as a series of tasks through Screen A and B. Most of the tasks are very intuitive, such as “Tap the Interests button at the top of the screen”.

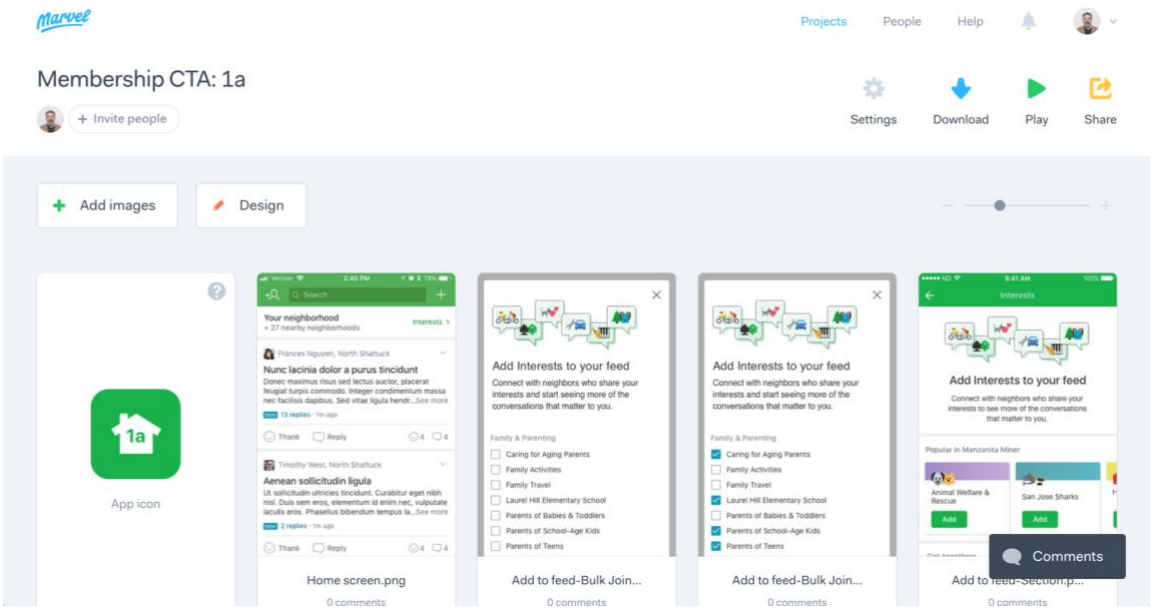

Figure 4: Prototypes of Task A

Following Task A, participants are then asked a few questions to gauge their “feel” of the feature. Keeping in mind that anything we ask may introduce new biases, we tried to the keep the tasks simple and the questions open-ended. Some of the questions that were asked include:

1. How would you describe Interests to a friend?
2. Does this seem different from how Nextdoor works now?
3. How did the description of Interests make you feel?
4. What do you expect to happen when you click <insert CTA>?
5. Is that appealing or not appealing to you?

While the designer on the project was responsible for writing the script, the team came up with a potential list of questions together. From the pool of questions, each one was

individually evaluated to identify the ones that were most likely going to assist in the synthesis of the study. Throughout the process, the team consistently questioned the purpose of each individual line in the script to avoid deviating from the focus of the study.

The script was initially written for the control group and then adapted to each of the variants. Usually, only a word or two needs to be changed across the scripts since the four flows only differed in their CTA options: Add to feed, Subscribe, Follow and Join.

Part 2 of the Test: Task B

This part of the user test focuses more on gaining insight into "what users say they do". Apart from just showing the users a single flow, each participant was taken to a screen with all four options at the end of the prototype.

Figure 5: All Four Variants Shown in Task B

They were then asked three questions, each directly corresponding to our three CTA goals:

1. Which version makes "Interests" the most appealing to you? Why?
2. Which version feels the most commitment-free? Why?
3. Which version gives you the clearest idea of what will happen when you click it? Why?

The method here incites users to reflect on their feelings towards the original variant relative to the new ones that are shown. It is essential to complement this part of the

study with results from the prototype tests in order to examine whether users "naturally feel" and "say they feel" a similar way. Often times, the two are misaligned, which is an important observation to point out in the synthesis of the study.

Adjusting for Biases

Although a conscious effort was already being made to remove any known biases from the test script, we forgot to consider how the order of tasks would impact the users' perception of all proceeding tasks. This is to say, two components of the test plan need to be adjusted to accommodate for the bias:

1. *The order of showing the bulk join flow and the section join flow* (Figure 3). To remove this bias, each of the four variant groups were further split into two subgroups, half of which saw Screen A first, the other half of which saw Screen B first.
2. *The order of Task A and Task B.* It should be noted that both the bulk join and section join flows fall under Task A entirely. Task B consists solely of the three questions related to the four variants which are shown together on the same screen. Here, the issue is that Task A acts as an anchor to the participant's response to the following three questions in Task B. Due to the fact that users have already been exposed to one of the variants, they may be more or less inclined to choose that variant when presented with all four options. Since it would require too many participants to conduct this test without bias, i.e. splitting up the test group even further, the stated bias will instead be accounted for when synthesizing results. The report will go into detail about it in a later section.

It is very common to not be aware of certain biases while designing the test plan. The key is to be conscious of the fact that this might occur and to continuously make an effort to question every decision that is made along the way [5]. As well, by gathering a variety people to look over the script, each individual can hopefully evaluate it under a different light, thus increasing the chance of catching a bias. Ideally, if time allows for it, the user test itself should be "tested" for biases, which can be done through releasing it first to a

small test group and then adjusting it accordingly before sending it out. While this strategy is not fool-proof, it is a still preventive technique and a good practice to introduce into a company's user testing process. [3]

4.3 Setting Up the Test

The usability tests would be conducted on a platform called usertesting.com since the design team at Nextdoor is already experienced in its product capabilities. As mentioned before, an unmoderated test approach was chosen to save time.

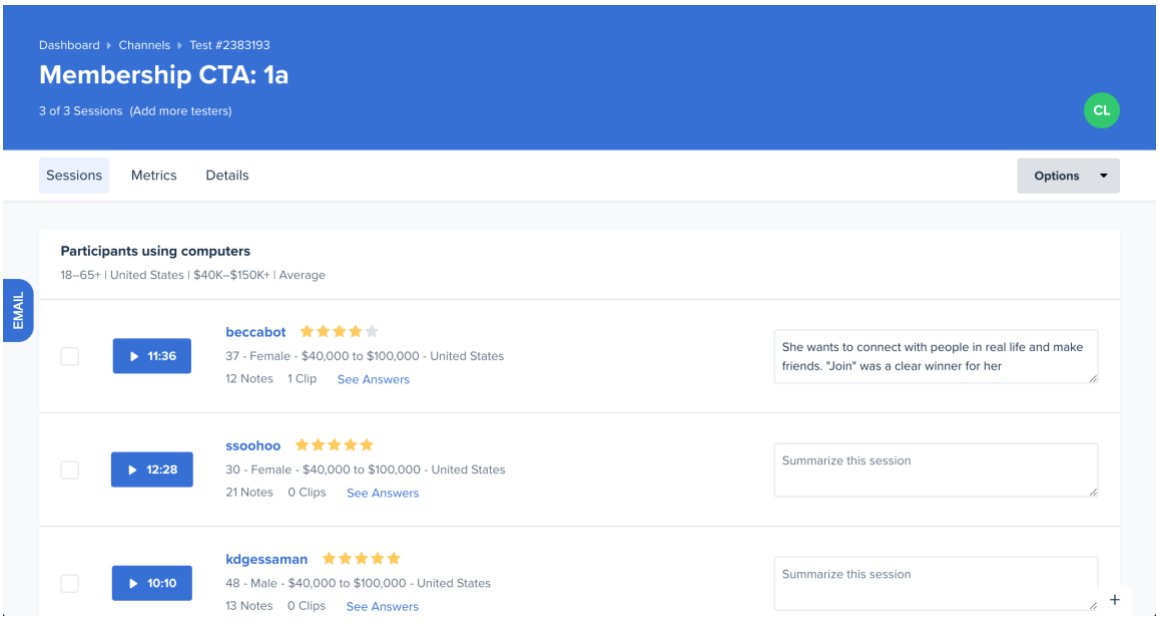

Figure 6: User Testing Platform

Selecting the Sample Size

Typically, it only takes 6 to 8 participants in each variant group to identify a usability pattern from the tests [6]. Less than 6, there would be insufficient information to synthesize any findings. Beyond that, the results become more or less redundant, with the exception of very complex and extensive user tests. While it is true that the larger the sample size, the more accurately the data pool models patterns in real life, the time and resources required are unnecessary for a simple test like the one that we are conducting.

[7] Thus, to optimize for time, the team decided to use the stated minimum of 6 users per variant, following this breakdown:

- Add to Feed
    - Group 1a (sees Bulk Join first)
    - Group 1b (sees Section first)
- Subscribe
    - Group 2a (sees Bulk Join first)
    - Group 2b (sees Section first)
- Follow
    - Group 3a (sees Bulk Join first)
    - Group 3b (sees Section first)
- Join
    - Group 4a (sees Bulk Join first)
    - Group 4b (sees Section first)

It should be noted that the previously mentioned biases are taken into consideration here.

Filtering Participants

Once the number of participants was agreed upon, the next step is to filter for users with specific characteristics on usertesting.com. The process encourages a response to the following question:

Who will help you answer your key questions and test your hypotheses?

To key here is to distinguish behaviors on Nextdoor (e.g. frequent inviter, new member, etc), behaviors outside of Nextdoor (e.g. loves eBay, organizes block parties, etc.), demographics, neighborhood types, as well as anything else specific that can be screened for during the recruitment process [1]. For our purposes, the CTA goals targeted users of Nextdoor that have not been previously exposed to the Interests feature. Thus, we filtered for these characteristics by asking three questions:

1. Which of the follow have you used: Pinterest, Instagram, Facebook, Snapchat, Nextdoor, LinkedIn?
    - *Must select* Nextdoor
2. Are you currently a member on Nextdoor?
    - (*Accept*) Yes
    - (*Reject*) No
3. How long have you been a member of Nextdoor: Less than one month, 1 - 3 Months, 4 - 12 Months, 1 - 2 Years, 3+ Years, None of the above?
    - (*Accept*) 4 - 12 Months, 1 - 2 Years, 3+ Years
    - (*Reject*) Less than one month, 1 - 3 Months, None of the above

The purpose of the first question is to obtain an understanding for the user's social media preference and proficiency level. The second question simply filters for active members on Nextdoor. While people might select Nextdoor in the first question, there is a possibility that they are a deactivated user or that they have previously used someone else's account. Finally, we are recruiting for participants who have not yet seen Interests on Nextdoor. Since all recent members who have joined within the past three months were automatically enrolled in the new Interests feature, we disqualified them in the selection criteria. The "None of the above" option simply shows again that they do not use Nextdoor, regardless of whether or not they selected it in the previous two questions.

In addition to the three questions, there were some mandatory demographic specifications required by the usability testing platform.

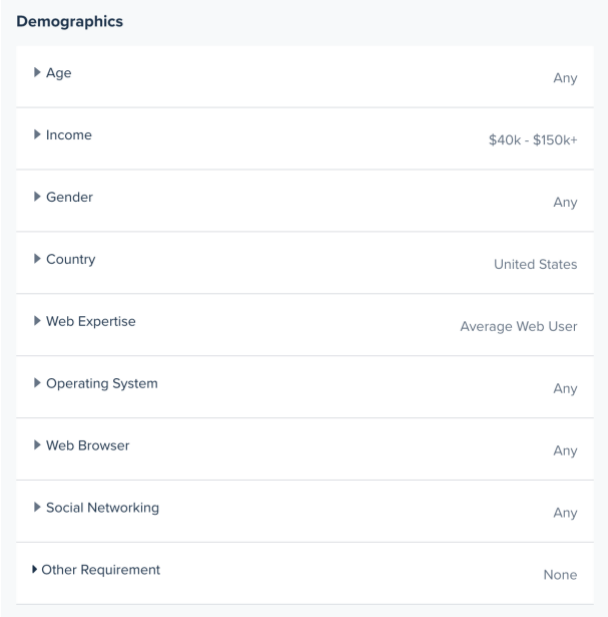

Figure 7: Target Audience Demographic Selection

As evident in Figure 7, a very broad audience was selected to test the prototypes as to accurately represent Nextdoor's user base.

4.4 Introducing the A/B Test

The A/B tests were later introduced into the test plan with the objective to support our qualitative analysis with quantitative data. This last part of the test plan focuses on the "how users behave" aspect of the study and complements the qualitative usability tests on usertesting.com.

The nature of the A/B test is very simple. Instead of recruiting users to test a prototype of our product, we released the four variants to a small portion of Nextdoor's actual user base and obtained metrics on user clicks.

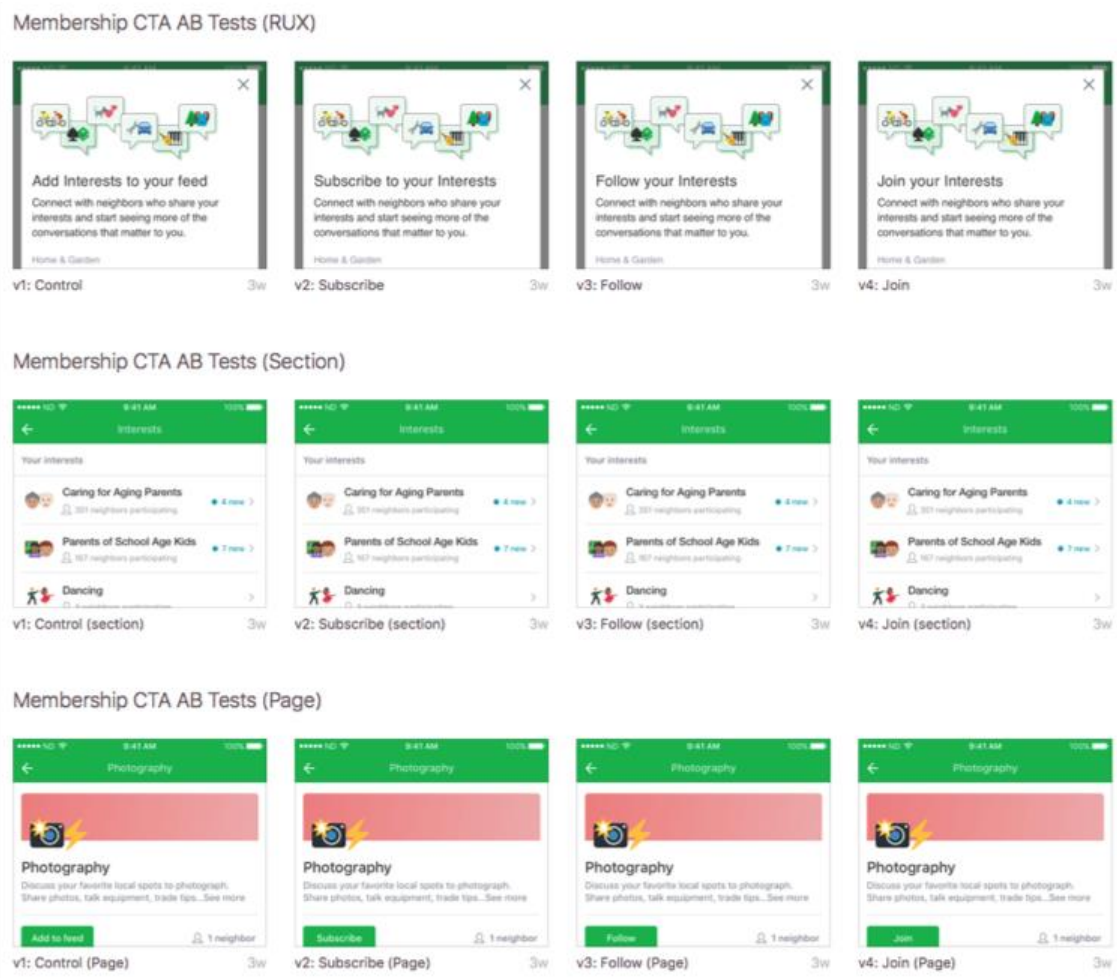

Figure 8: A/B Test Variants

The same audience selection criteria that was used on usertesting.com was correspondingly applied to this user test group. The only difference for the A/B tests is that a computer algorithm selects by random the variant that the chosen users get to see. There is nothing qualitative about this test, and results only provide a value pertaining to the membership engagement of each variant [3]. This is to say, the A/B tests reflect exactly how users react to the Interests feature (they either opt-in or they do not), rather than how they "think" they will use it.

5

Analysis

5.1 Synthesizing the Usability Test

The data from the conducted usability tests can be analysed from many different angles. To start off, notes were made beside all test recording videos to capture any insights on the problem that are hoping to solve. From this, the observations can then be translated into a more quantitative summary.

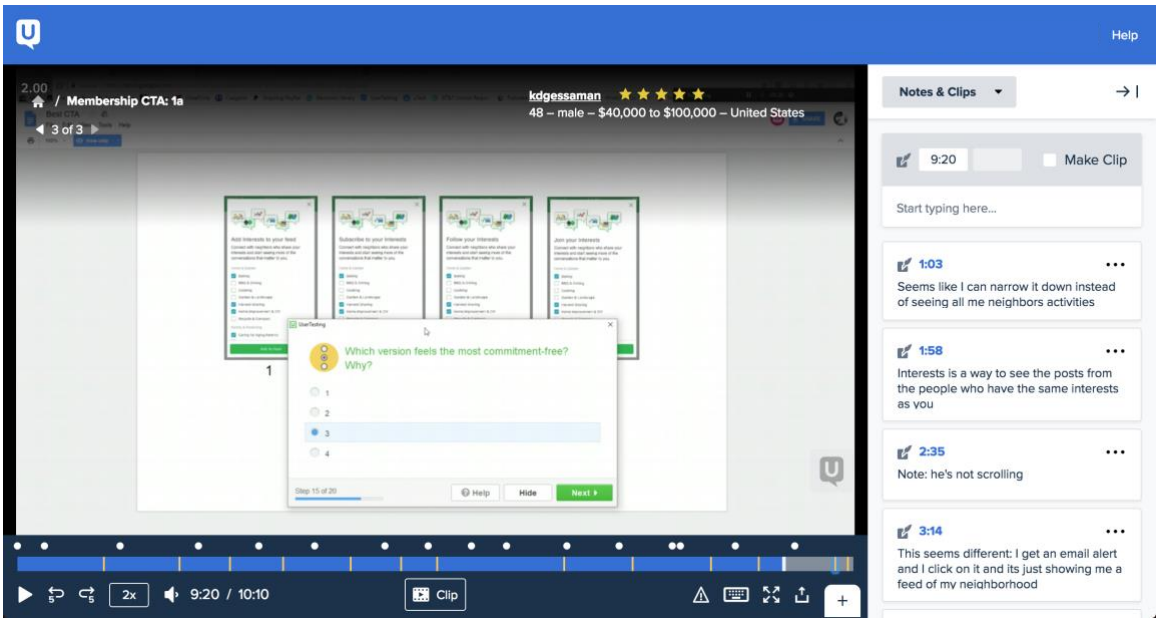

Figure 9: Making Notes on Usability Tests

Synthesizing Task A

In the following Table, the notes from Task A are summarized and organized by variant. The synthesis above approaches the test results by rating each participant on a scale from 1 to 5, where 1 represents a poor understanding of Interests and 5 represents an accurate understanding. The results from each variant group were then totalled to obtain 67, 66, 81 and 77 for Add to Feed, Subscribe, Follow and Join respectively. From this analysis, we

are able to gauge the users' understanding of how Interests works, as well as their perception of its value in the application.

Table 3: Synthesis for Task A

What did they see first?		How would you describe it	rating	How might people use it	rating	What do you exect <CTA> to do	rating	Is it appealing or not?	Y/N	TOTALS
Add to feed	1a		1-5		1-5		1-5			
	beccabot	It would be about connecting with people you normally would	3	It's a little vague about how people would use interests like a	3	Not entirely sure what to expect. She want's to know if it's use	2	It's very appealing - I'm pretty isolated and don't know many p	Y	
	soohoo	Interests is a way to filter your Nextdoor feed so that you see	4	This is different from how it works now because it's curated n	3	I would see posts from other people who checked the same t	3	This is appealing because I get messages about a trending p	Y	
	kdgessaman	Interests is a way to see the posts from the people who have	3	Seems like I can narrow it down instead of seeing all me neig	3	I expect to see posts from people who share these interests	3	I think it's appealing. I like to interact with people locally - abo	Y	
	1b									
	Delalcon	In the neighborhood app there's a section where you can ad	4	I imagine you can set up events, clubs, there'd be a calendar	4	You'll have access to little chatrooms	3	It's a good idea - it's great that ND wants you to get out and n	N	
	alonb1977	Interests are topics and things that you feel that you like or y	5	Basically we're making nextdoor a little bit more social - prob	4	Automatically my feed is going to have content related to thos	4	I must admit it is appealing to me because there a lot of times	Y	
	abanta87	Its a way to find like-minded people with similar hobbies that	4	I wouldn't click all the things I'm interested in if it doesn't seen	4	Add makes me think it's adding a tag to my profile page, but i	3	This turns it into much more of a social network. Right now m	Y	
			23		21		18		5	67
Subscribe	2a									
	Jmm262	Nextdoor is more of a safety site or for getting help. This is try	2	I guess this would hook us up ways to get together with your	3	Not sure what I'm subscribing to - maybe its just a group. I as	3	I don't find it appealing for nextdoor - it seems separate, like "	N	
	andrea8090	It's a way to make the more relevant content more prominent	4	You wouldn't have to do anything besides subscribe and ther	4	My biggest question is how do my interests play into this? do	2	This is a great step in the right direction in making the app mo	Y	
	Reneasc	Drills down to something specific so that you're not scrolling t	3	Maybe the weekend is coming up and I want to look at all the	2	How does it work? I'm going to get just that area in which I'm	2	Kind of sort of - I'm not sure if there will be two tabs, one listin	N	
	2b									
	ca2fa	Interest-based way to connect with neighbors instead of just t	4	I don't know if its just going to be my neighborhood or the res	4	My guess is that I'd now get a new newsfeed subsection ONL	4	It's appealing, but I'd like to know more about it. How are the i	Y	
	mib2n	To me Interests is a way to subscribe to groups in the area - t	5	More social element to connect with people who share the sa	5	Expect subscribing to a forum with other people in the area to	4	The fact that you can opt in and subscribe is appealing becau	Y	
	darkbulb	I like the concept - the word itself (Interests) is really descript	4	**That's what I'm struggling with a little - Struggling with w**	3	Thinks of these as groups and wants to know how many mer	4	I like the first part ("Connect with neighbors who share your in	Y	
			22		21		19		4	66
Follow	3a									
	Jelleo	People who have the same interests in one thing... I don't kn	4	Makes me feel like I'll get to know my neighbors a little bit mo	4	When I click follow, I expect to be added to some of these inte	4	It's appealing because its kind of familiar. it seems to be kind	Y	
	epathj	Topics or subjects that are part of your life for different reasor	5	These would be part of your feeds on your homepage. Maybe	5	That's now going to be part of my profile and that these intere	5	Takes the app into being more of what a neighbor is all about	Y	
	RJTester729	People that may be in a similar situation to you that you want	5	I'd expect to meet up with people and have a group page or g	5	Follow would put me into a group page with other people with	5	Its appealing, sounds like a really smart idea. This is really gr	Y	
	3b									
	bbbird	Interests - perfect word for it. Arrange get-togethers and fun v	5	This is different from how I use it now, which is to share wher	4	Expect to follow people who are interested in those things - n	2	I like what I see - I'm anxious to explore this and see what kin	Y	
	rcollier87	It's like twitter where I follow someone, but more like a categ	5	It's coming a bit more like meetup as well - another way of bu	5	I didn't even read the description before but I knew just by the	5	Appealing: Motorcycles would be appealing. Riding with othe	Y	
	DFrank	Its a way of sorting conversations between people based on	4	I'd expect to have a special feed of those topics coming throu	4	Makes me feel like facebook. I'm not sure if I like that or not.	N	I'm not sure if its appealing, but I'm not sure exactly how its g	N	
			28		27		21		5	81
Join	4a									
	SCasey4444	I'd explain: There's a place now on ND where you can conne	4	I'm not sure if this is about meeting up or just just conversatio	4	I'm not sure what to expect for how I actually get connected?	2	This is appealing though - absolutely	Y	
	tx5smith	I can join groups with the same interests it looks like. They're	5	Seems like another website I have to be compartmentalized i	3	Join means that it's just going to add stuff to my newsfeed	5	Joining interests is not appealing to me	N	
	hawksnst	An area you could see other comment from people who are i	4	This is such a departure from what I'm used to seeing. I gene	3	I expect nothing to happen at first, and then if something is p	5	I'm not sure if this is appealing - you have to do it on linked in	N	
	4b									
	PoeDiggity	Any kind of mutual interest you could have with your neighbo	3	If I wanted to have a get-together in the neighborhood would	4	Join will add me to that group so I can view all the threads an	5	Yes - looking forward to getting to know my neighbors a little t	Y	
	lk	Would describe it as a way to FOLLOW topics youre interest	4	Feels like it could be a way to build relationships with people	4	I expect it to join me in that interest and take me to the interst	4	Yes thats appealing if it takes me directly to the interest page	Y	
	eyeteachart	Kind of like closed groups on facebook, limited to just your cc	5	You'd be able to have surveys or forums right there in the nei	5	I expect to see a description of the group, who's in it, an annc	4	**Clip** interests is appealing to me - it streamline what I see.	Y	
			25		23		25		4	77

In Table 3, the variants Follow (in green) and Join (in yellow) won by a significant amount over Add to Feed and Subscribe. However, these observations are simply insights, and conclusions cannot be made without analyzing the results in conjunction with the other two tests. The mentioned synthesis process is also extremely subjective and may be biased based on the person that is evaluating. Incorporating another pair of eyes on the synthesis can definitely be beneficial. However, it is important to note that all evaluators must look over all 4 test variants, rather than distributing the variants across different people. In doing so, at least the subjectiveness due to potential evaluator bias remains consistent throughout each individual analysis. Later, these values can then be averaged to obtain a less biased result.

Synthesizing Task B

A similar procedure was taken to analyse the results from Task B, where all 4 variants were displayed on the same screen. In a spreadsheet, the answers from each participant group was recorded, and a number was tallied for the 4 variants based on which CTA was found to be the most appealing, commitment-free, and clear to users. As mentioned

before, the bias that resulted from the order of tasks could not be removed. Thus, a slightly more extensive approach was taken to synthesizing the findings.

Table 4: Unbiased (Left) and Biased (Right) Synthesis for Task B

What did they see first?		1 Add to feed	2 Subscribe	3 Follow	4 Join
Add to feed	1a				
	Appealing	0	0	2	1
	Commitment free	1	0	2	0
	Clearest expectations	1	0	1	1
	1b				
	Appealing	2	1	0	0
	Commitment free	2	0	1	0
	Clearest expectations	1	1	1	0
Subscribe	2a				
	Appealing	0	1	0	2
	Commitment free	1	0	1	1
	Clearest expectations	2	0	0	1
	2b				
	Appealing	1	2	0	0
	Commitment free	1	0	2	0
	Clearest expectations	1	2	0	0
Follow	3a				
	Appealing	3	0	0	0
	Commitment free	3	0	0	0
	Clearest expectations	2	0	0	1
	3b				
	Appealing	0	0	1	2
	Commitment free	0	0	2	1
	Clearest expectations	2	0	0	1
Join	4a				
	Appealing	1	0	1	1
	Commitment free	2	0	1	0
	Clearest expectations	1	0	1	1
	4b				
	Appealing	1	0	2	0
	Commitment free	1	1	1	0
	Clearest expectations	3	0	0	0
		32	8	19	13

What did they see first?		1 Add to feed	2 Subscribe	3 Follow	4 Join
Add to feed	1a				
	Appealing	0	0	2	1
	Commitment free	1	0	2	0
	Clearest expectations	1	0	1	1
	1b				
	Appealing	2	1	0	0
	Commitment free	2	0	1	0
	Clearest expectations	1	1	1	0
		7	2	7	2
Subscribe	2a				
	Appealing	0	1	0	2
	Commitment free	1	0	1	1
	Clearest expectations	2	0	0	1
	2b				
	Appealing	1	2	0	0
	Commitment free	1	0	2	0
	Clearest expectations	1	2	0	0
		6	5	3	4
Follow	3a				
	Appealing	3	0	0	0
	Commitment free	3	0	0	0
	Clearest expectations	2	0	0	1
	3b				
	Appealing	0	0	1	2
	Commitment free	0	0	2	1
	Clearest expectations	2	0	0	1
		10	0	3	5
Join	4a				
	Appealing	1	0	1	1
	Commitment free	2	0	1	0
	Clearest expectations	1	0	1	1
	4b				
	Appealing	1	0	2	0
	Commitment free	1	1	1	0
	Clearest expectations	3	0	0	0
		9	1	6	2

Without considering the impact due to bias, the table on the left tallies the results for each variant in Task B. The table on the right also sums up the values but keeps them categorized according to the flow that users were taken through first (1 of 4 variants in Task A). This helps to with analyzing the contribution of bias towards the overall number. An example of the bias taking affect can be observed: 5 out of 8 points (62.5%) attributed to the “Subscribe” column comes from users who initially saw the Subscribe screen in Task A.

It should be noted here that the findings from Task B deviate a bit from that of Task A. The Add to Feed CTA won by a fair amount in this synthesis, in comparison to Follow and Join which performed the best in Task A.

5.2 Synthesizing the A/B Test

The A/B test naturally produced quantitative metrics for analysis. The figure below portrays the findings in a statistical format based on the CTA click-through rate. While

the statistical depiction of the data itself seems very complicated, the most important takeaway can be understood by simply looking at the length of the green bar, which quantifies improvements to the click rate.

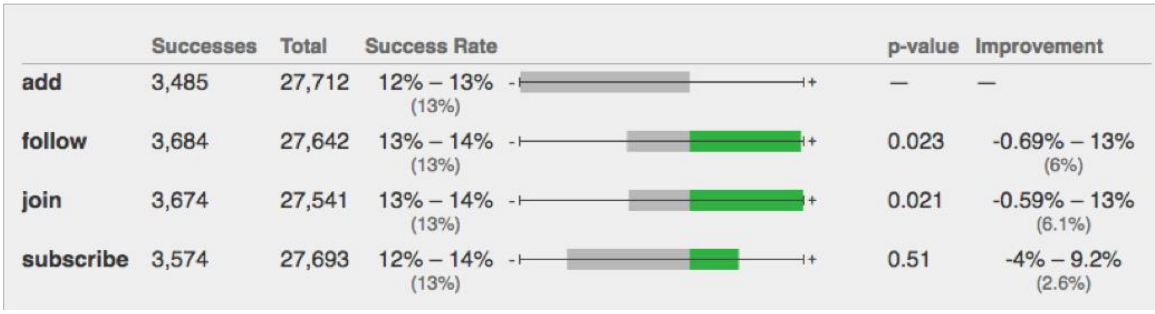

Figure 10: Synthesis for A/B Test

The reason that the results are represented as a bar rather than one single metric is due to the fact that a "sample" was taken instead of the entire population. Thus, the bar allows the results to fall in a range of values to accommodate for the statistical inaccuracy. [7] Based on this visual interpretation, Follow and Join both performed better than the other two CTAs. These results align with that of Task A but not Task B, whose clear winner was "Add to Feed".

6

Discussion of Results

6.1 Usability Test Results

Add to Feed

People like the mechanical, straightforward nature of this wording. They understand (or at least think they understand) how it affects the product. It however confuses more people than any other variant in terms of how it actually works. Many users also mentioned the appeal of being able to filter what they see in their feed.

Subscribe

To users, this feels like an older paradigm, where they must sign up to receive broadcasts and pay a fee. It was the least performing CTA in every qualitative analysis.

Follow

Due to their experiences on other platforms, people like that this is a term they are already comfortable with. It feels commitment-free and as if one can be a passive observer of content, which happens to be appealing to most. This wording does not appeal to those who wish that the feature would help them make more connections.

Join

The people who liked Join seemed to be the most excited about the prospect of the Interests feature. They thought that it sounds friendlier and more "neighborly" while still capturing the fact that they would not just be consuming, but also contributing content. These people talked about joining activities, creating events and meeting new people, as opposed to simply viewing relevant content in their feed. They understood how the feature affects the product as well.

6.2 A/B Test Results

In the A/B test, Follow and Join both outperformed by 6% in terms of subscribers to Interests. This however did not impact the number of subscriptions generated by these subscribers, as all 4 variants fell between an average value of 9.5 to 10 subscriptions per subscriber. Further testing is required to obtain conclusive results.

6.3 Key Insights

For the most part, the results from the qualitative usability tests and quantitative A/B test aligned with one another. Follow and Join were the winning CTA in both cases, with the exception of Task B. In Task B, while people thought that Add to Feed was overall the most commitment-free, light-weight, and clear (obtaining a significantly higher score than the rest of the CTA options), this variant did not actually receive a lot of click-through in the A/B test. The variants that were most successful in the A/B test were instead Join and Follow. This is consistent with how users felt in Task A, where Join and Follow scored approximately 10 points higher than the other two variants. In all of these tests, subscribe was the worst performing CTA. As our findings supported the fact that people often act differently from how they say they would, we identified/determined the results from Task B to be inconclusive, which left us to decide between the Follow and Join CTA.

In order to justify choosing one CTA over the other, it is important to think through what we want people to do on Nextdoor, or specifically Interests. The purpose of Interests is "to gather the community together for activities and events that otherwise would not have happened". In this sense, the goal is not simply to have people consume content on Nextdoor, but also to encourage subscribers to create content themselves.

Strategic Decision

Comparing the two options, Join is strategically better since it aligns more closely with Interests as a feature. From analysing the usability tests, people associated Follow with passively consuming content in their feed. By contrast, users who found Join to be appealing were more inclined to engage with their community as they felt that the feature

would help them connect with real people who they can relate to. Thus, with the objective to drive membership engagement, the final decision was to move forward with "Join" as the new Interests CTA.

7

Conclusion

From executing the usability tests, we were able to take the problem at hand and analyse it from many different angles to make it a more comprehensive analysis. In the first part of the study, we investigated how users "feel" about the Interests feature by asking them to perform a series tasks on a Marvel prototype. In the second part of the study, users were shown all 4 variations of Interests and were encouraged to "say" what they thought about the feature. Finally, the last part of the analysis seeks to understand how users actually "behave" on the platform based on real-life metrics. Referring back to the objective of the study, which was to identify the best CTA in terms of how commitment-free, light-weight, and clear it appears to be, it was concluded that "Join" and "Follow" performed the best in both the qualitative and quantitative analyses. Due to the lack of an obvious winner, "Join" was strategically chosen as it best aligns with how we hope that users would interact with the Interests feature.

8

Recommendations

To gain more insight from usability testing, a good approach is to seek ways to obtain more accurate findings while digging deeper into the problem. Here, two major recommendations can help achieve that:

- In all the usability tests, we made the assumption that the qualities "appealing", "commitment-free" and "clearest expectations" are valued equally by our users. Thus, we proceeded to distribute the weight of their impact equally in the synthesis process. However, this might not actually reflect reality, because users may care more about certain characteristics over others. In addition, the tests in this study do not seek to analyse the effect of each individual quality on membership engagement. While it is indeed important for the CTA to feel "commitment-free", it is possible that there are negative correlations associated with it, such as a reduction in content creation. To further understand the problem scope, another test should be conducted to evaluate our users' perception of these qualities in more depth, in addition to how it might factor into engagement metrics. Such insight should be very beneficial in supporting our findings, but was simply not a focus of the current usability study.
- Based on where we landed in the study, both Follow and Join came in close as winners. However, the team made a strategic decision to implement the Join CTA. Moving forwards, a more conclusive study would be to send out another A/B test that compares the behaviour of only the Join and Follow CTAs. Since we are assuming that Join will perform better, it might be a good idea to start with an 80/20 test, where 80% of users see Join and 20% see Follow.

The following are a few additional recommendations that were mentioned throughout the report, but were unable to be integrated in the study:

- We should "test" the user tests prior to release to observe for any unaccounted-for biases.

- We should inform decision-making with a greater variety of research methods – qualitative, quantitative, behavioural and attitudinal. In this study, we only touched upon two user testing strategies due to lack of time and resources.
- We should aim to further reduce biases from our study by recruiting people with different backgrounds to review the test plan and to help synthesize results.

By implementing these recommendations, the findings from this study would be better supported from a variety of angles, thereby allowing it to become a more thorough analysis.

References

[1] "What is User Research?," *The Interaction Design Foundation*. [Online]. Available: https://www.interaction-design.org/literature/topics/user-research. [Accessed: 10-May-2018].

[2] R. Budiu, "Quantitative vs. Qualitative Usability Testing," *Nielsen Norman Group*, 01-Oct-2017. [Online]. Available: https://www.nngroup.com/articles/quant-vs-qual/. [Accessed: 11-May-2018].

[3] C. Rohrer, "When to Use Which User-Experience Research Methods," *Nielsen Norman Group*, 12-Oct-2014. [Online]. Available: https://www.nngroup.com/articles/which-ux-research-methods/. [Accessed: 11-May-2018].

[4] C. Wickens, *An Introduction to Human Factors Engineering.* 2nd Edition, New Jersey: Pearson Education Inc., 2004

[5] W. Lidwell, K. Holden, and J. Butler, *The Universal Principles of Design*. Beverly, MA: Rockport Publishers, Inc., 2015.

[6] J. Nielsen, "Why You Only Need to Test with 5 Users," *Nielsen Norman Group*, 19-Mar-2000. [Online]. Available: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/. [Accessed: 12-May-2018].

[7] "What Is Data Analysis?," *Data Collection Techniques | CYFAR*. [Online]. Available: https://cyfar.org/ilm_6_2. [Accessed: 13-May-2018].