LORE
Contagious health

User Testing LoreBot 2024-2025

LoreBot Chat History UI Version 1.0

“Preference” Usability Test

October 16, 2024 (Test Date)

Full Test Results

Testing Methodology

A **Preference Test** is a testing method where users choose options based on **subjective preference** on like visual appeal, clarity or initial impression.

It answers the question: *“If users are given a grouping of past conversations, how do they expect them to be grouped?”*

Test Plan and Purpose

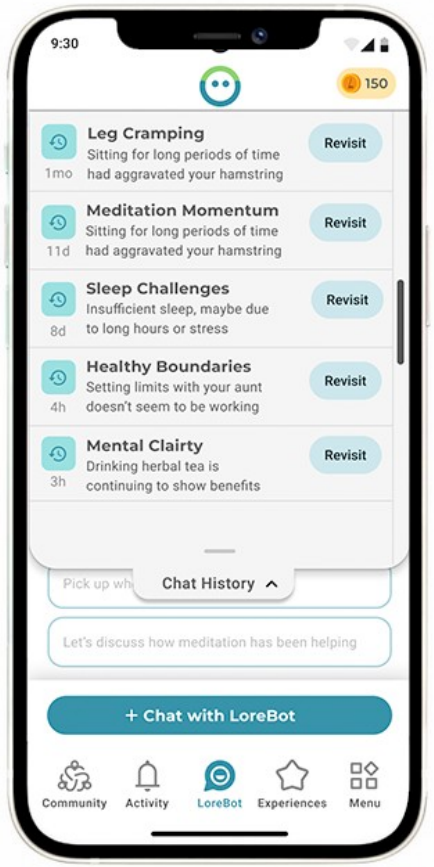
Purpose
Understand how people might expect past conversations with LoreBot to be shown

Method
Pick the favorite or helpful grouping out of three options

Participants
26 participants most based in the US and UK, not familiar with the UI, nor Lore

Duration
Writing + deploying test: 1h
Response time: 1h
Synthesis + Analysis: 2h (no deck)
Est. Test Cost: $50 (to vendor)

Results Overview



For reference:
Chat History Top Panel

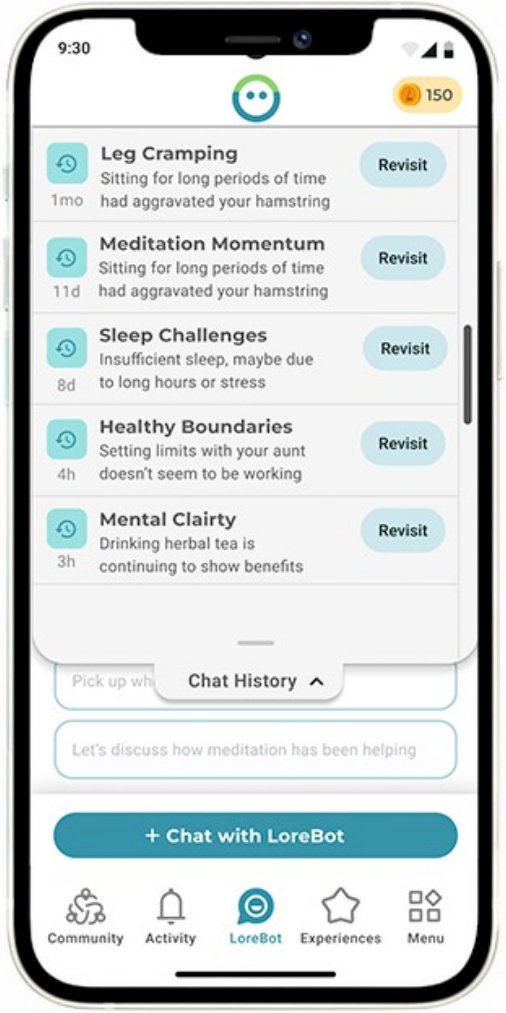
The purpose is to understand how users might think of their personal past, and how they might visualize previous conversations or actions within LoreBot.

Pre-user test, our squad was firmly divided over by topic or date.

Option	Grouping Schema	Example	Benefit
A	By Session	"Leg Cramps", "Leg Cramps"	Easier to do for us 😐
B	By Topic	"Difficulty Running"	Two-word summary 😐
C	By Day / Time	Yesterday, Tuesday	Potentially relevant for recall 😐
D	***The fourth version—a discovery while the squad was reviewing the findings here in the deck***	**[Variable based on the user's frequency, topics, sessions]**	**Highly-personalized*** 🙂

* *Could go even further to learn how one's mental model and real use align (deep task, deep solution)*

Option A: Display by Session



Test Option A

Overview of Option A

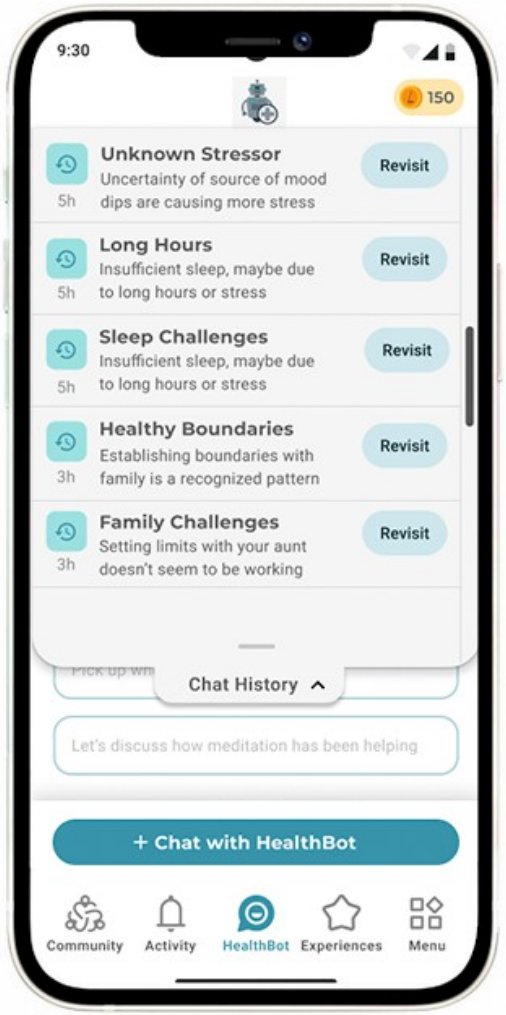
LoreBot currently recognizes a "session" as 40 minutes.

It's a baseline measurement arbitrarily used to attach events to (like resetting Conversation Pills or counting traffic to the app.)

It's easiest at the moment to make things "sessions" however the users will not know how it's grouped / understand the logic.

* *In the study, since the topics are prefabricated, participants may relate to the topics more than the format*

Option B: Display by Topic



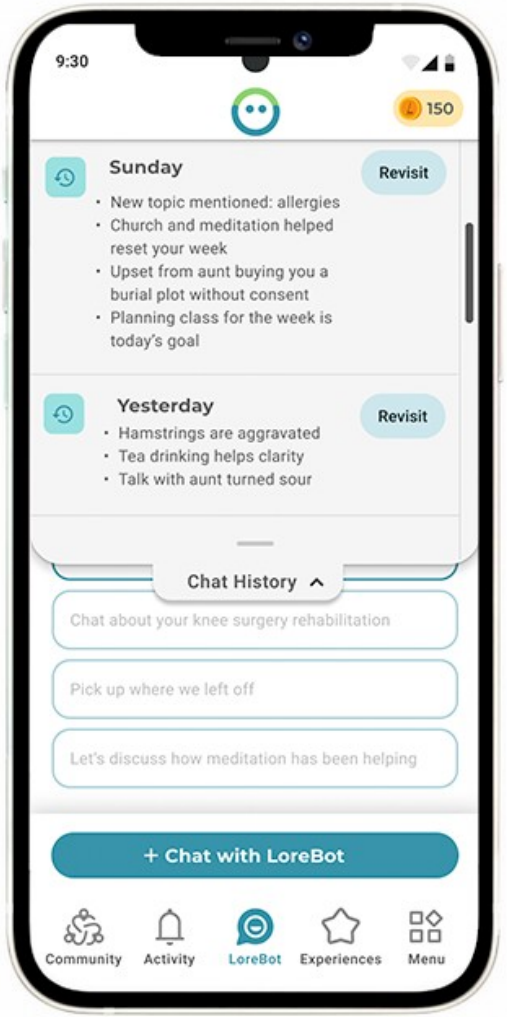
Test Option B

Overview of Option B

A series of 'appropriate' AI-gen condensation of groupings determined by 'mining' and recombining into logical groupings for the user.

* *In the study, since the topics are prefabricated, participants may relate to the topics more than the format*

Option C: Display by Time



Test Option C

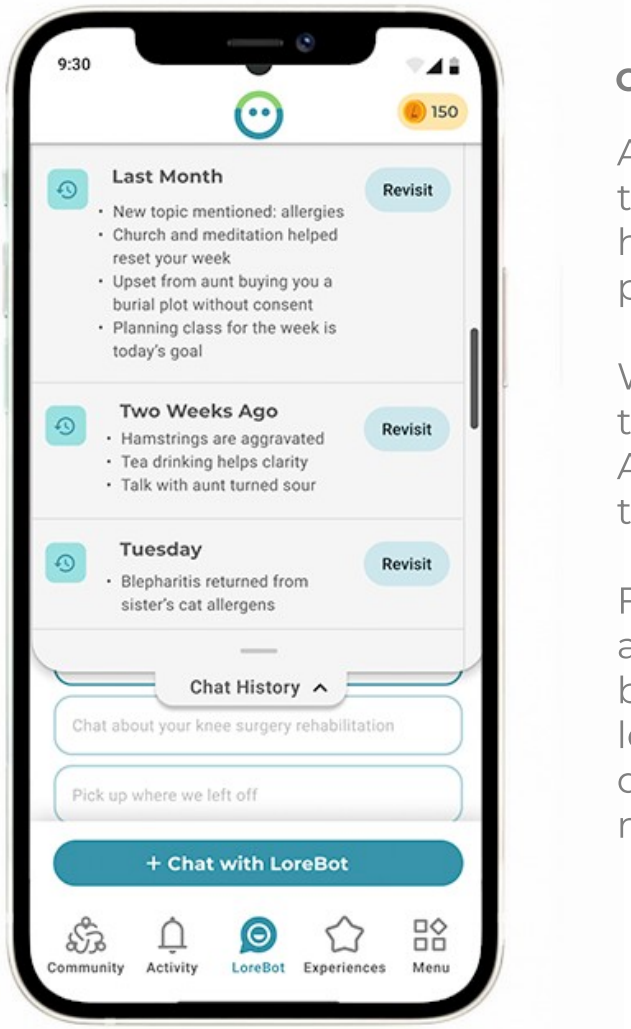
Overview of Option C

Using weeks, days as groupings.
As the user moves forward in time, groupings become wider.

For example last Sunday and Monday, might merge into "last week". If someone is a frequent daily user, they would see groupings happen more often.

* *In the study, since the topics are prefabricated, participants may relate to the topics more than the format*

Discovery of Option D: Custom Tailored Groupings



Test Option D

Overview of Option D

After bringing the test results back to the squad, we saw that the division we had between topic and time was also present in the 26 participants' answers.

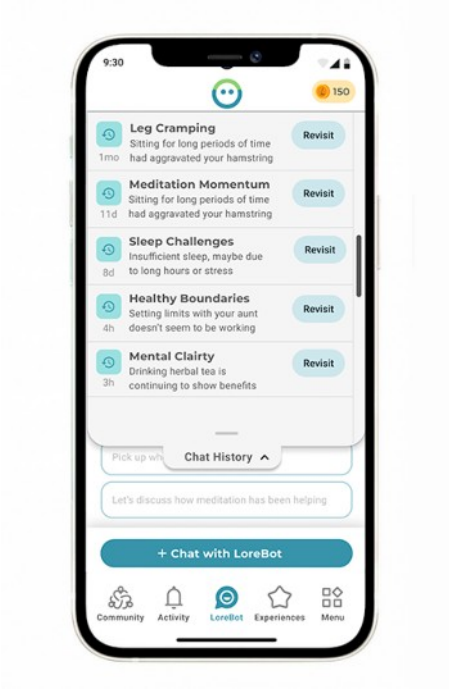
We talked about the idea of putting the data first having that reviewed by AI, and then returned to the user in their usage format.

For example, a daily user who jumps around to multiple topics we, might break it out more by topics. But really leaving it to the content and time to display appropriate for their mental model and use.

Participant Profiles + Results

Participants

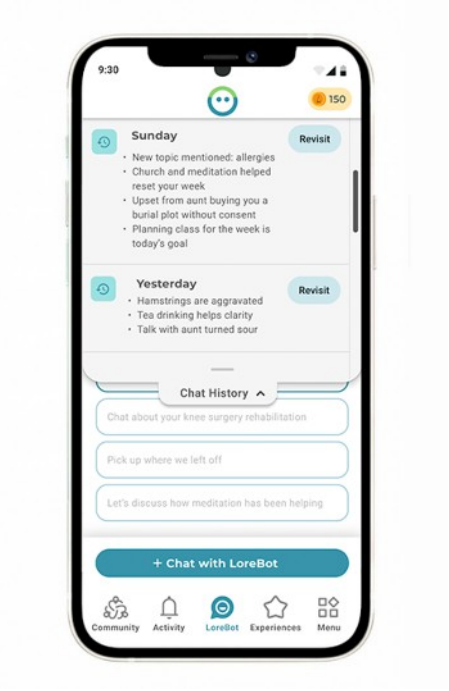
- 26 Teachers– elementary through high school
- $40-100k
- Mostly US, then UK, GER
- Ages 25-59



Option A: Display by Topic

38%

10/26 Participants



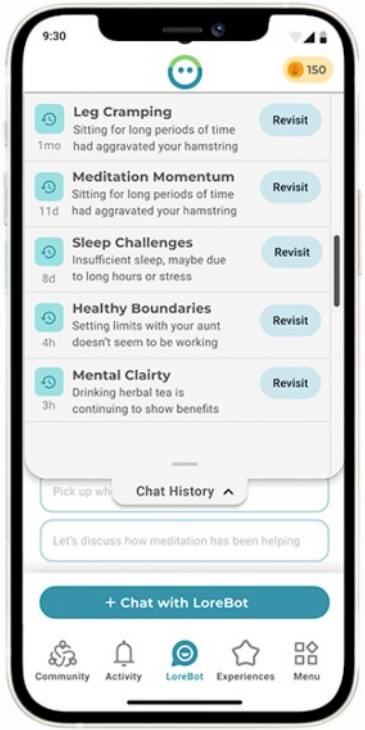
Option C: Display by Time

42%

11/26 Participants

There was a similar split in our squad.

Participant Quotes



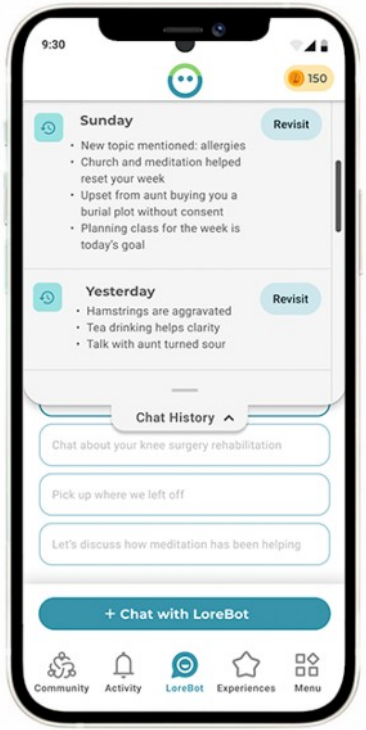
I like the format of A the best. The headings it used for each topic helps to focus on on specific socio emotional skills.

The topics is more helpful for me to reflect on specific topics

Option A: Display by Topic

38%

10/26 Participants



This is the most condensed and easiest to digest. The other options were overwhelming.

I like how it explained exactly what the chat was about with bulleted summary, and gave the exact day.

Option C: Display by Time

42%

11/26 Participants

This feedback led our squad to a discussion to be able to merge two into an adaptive solution.

Outcomes + Next Steps

User Testing Yielded an Immediate Result

Sharing this exercise with the squad led to a conversation for a blended/hybrid version based on the users' data structure.

Scope of Research

In this case, the two avenues of user testing showed similar results and created a better understanding than just one point-of-view (ours)

Real Content

Having real use, real scenarios, real blurbs users are familiar with will be a great way to learn, iterate this idea further

LoreBot Landing Page UI Version 1.1

“First Click” Usability Test (Pilot)

June 6, 2025 (Test Date)

Testing Methodology

A **First Click Test** is a usability testing method used to evaluate **where participants instinctively click first** on a screen to complete a task.

It answers the question: *“If users are given a goal, where do they think they should begin?”*

Test Plan and Purpose

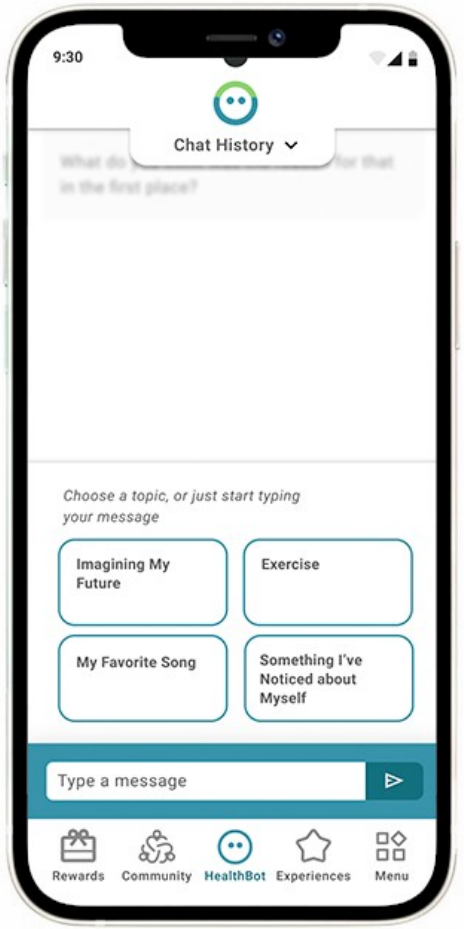
Purpose
Evaluate user behavior across four common LoreBot scenarios

Method
First-click test using an identical screen across 4 scenarios

Participants
7 participants based in the US, not familiar with the UI, nor Lore

Duration
Writing + deploying test: 25m
Response time: 40m
Synthesis and Analysis: 4h
Est. Test Cost: $50 (to vendor)

Results Overview



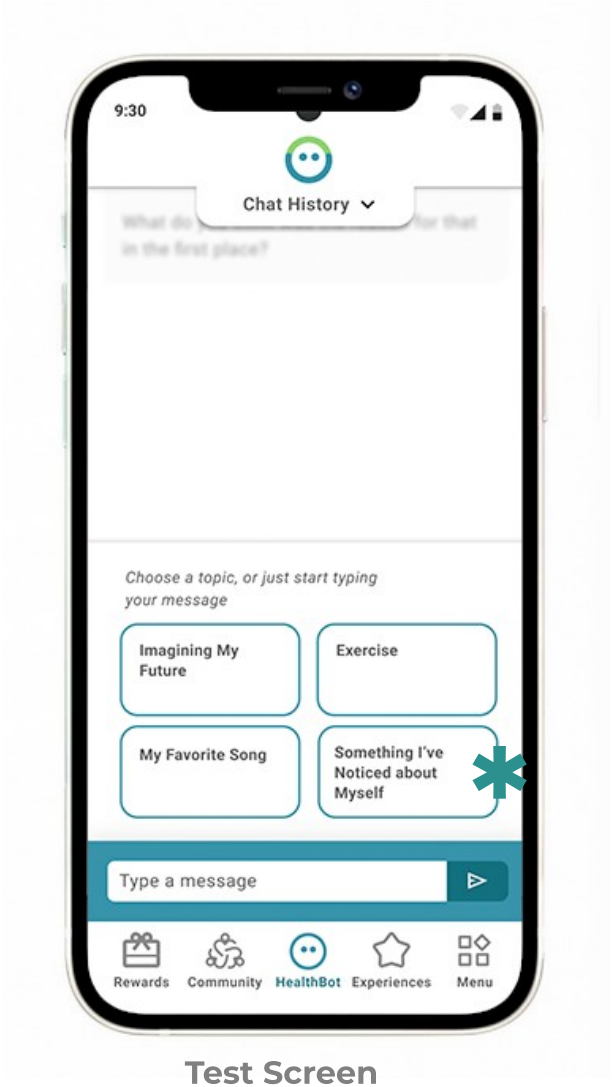
Test Screen

The purpose and actions expected from the users was generally well understood, even in a static test design.

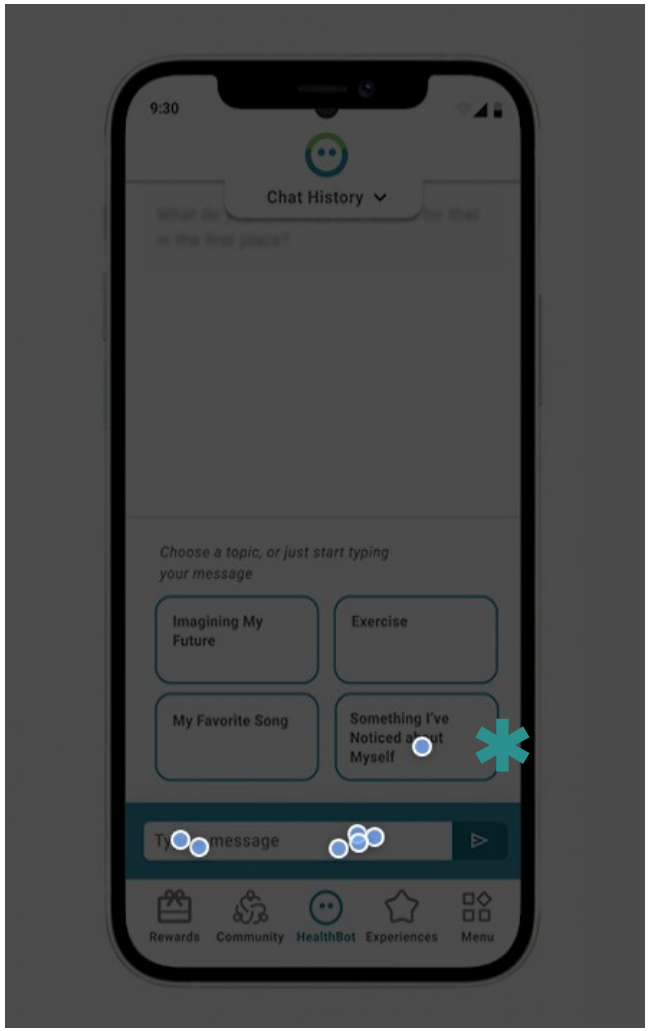
Features explained as, or tied to, a duration were less clear.

Scenarios	UI Element	Takeaway
New Topic	LoreBot Text Field	Understood as open topic area 🙂
Unsure of a topic	Topic Pills	Pills a strong draw 🙂
Recall old conversation	Chat History	Very clear in label / function 🙂
Recall a very recent conversation	Chat Transcript	Cloudy, but very dialed down visually in test design 😐

User Scenario 1
Ready to chat on the user's own topic



Test Screen



Users' Tap Pattern (7 total)

Results

6 of 7 tapped "LoreBot Text Entry" UI

Strong User Confidence (avg. 5 / 5)

Notes + Future Steps

May have cleared the late 2024 issue of users thinking a pill above was the subject to talk about.

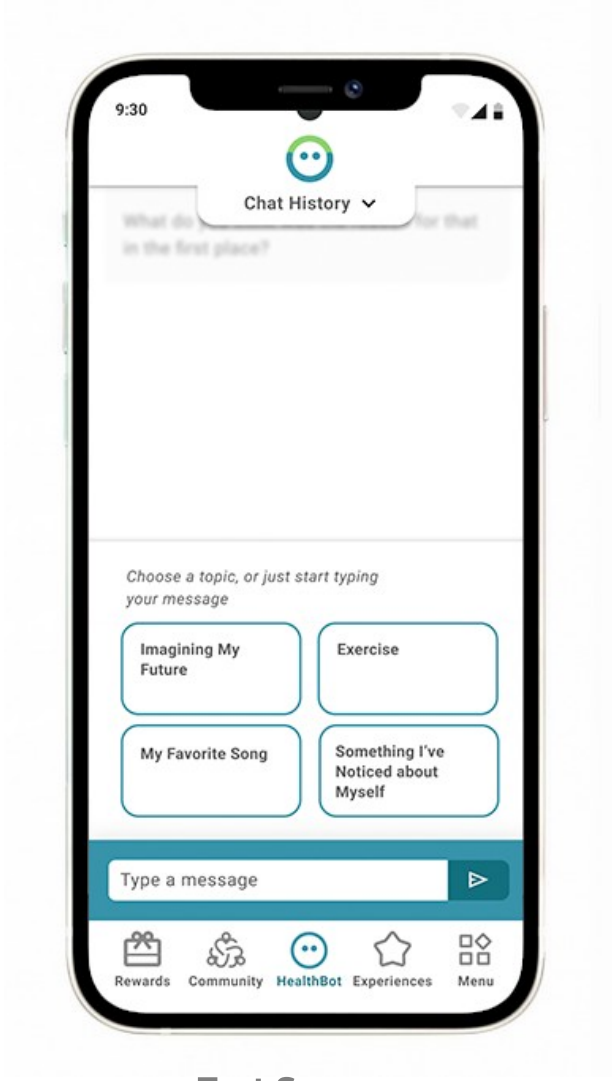
Will continue to learn about this / watch for that.

Experiment with different support language in the text field

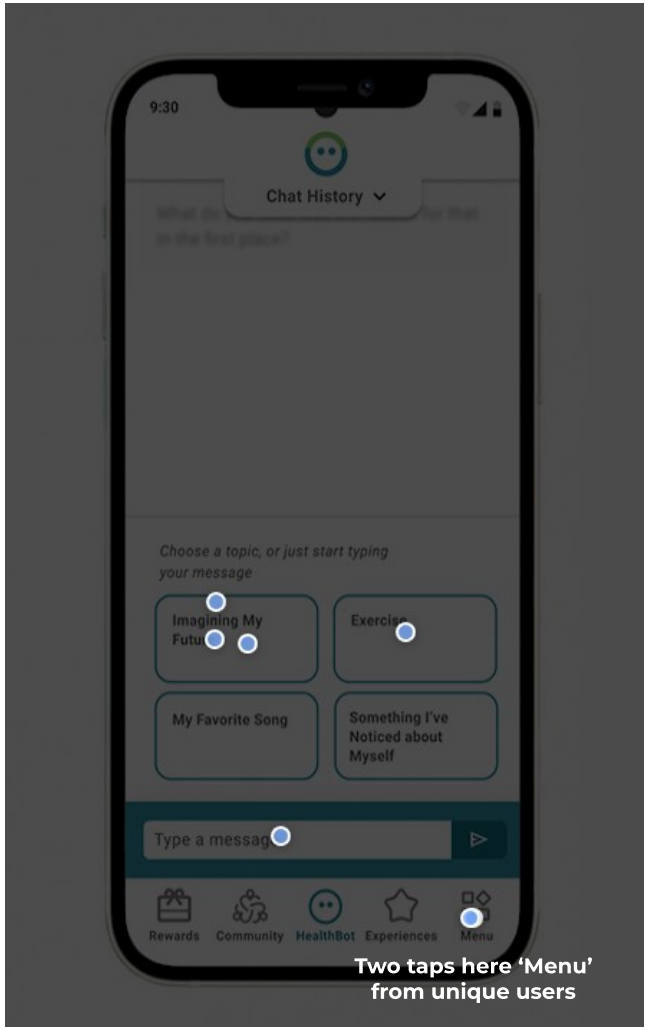
* The outlier tapped the pill "Something I Noticed about Myself"—which is valid!

Triangulate this data against future tracking, and in-the-field observation

User Scenario 2
Ready to chat but no specific topic



Test Screen



Users' Tap Pattern (7 total)

Results

4 of 7 tapped "Conversation Starter" UI

Moderate to Strong User Confidence (avg. 4 / 5)

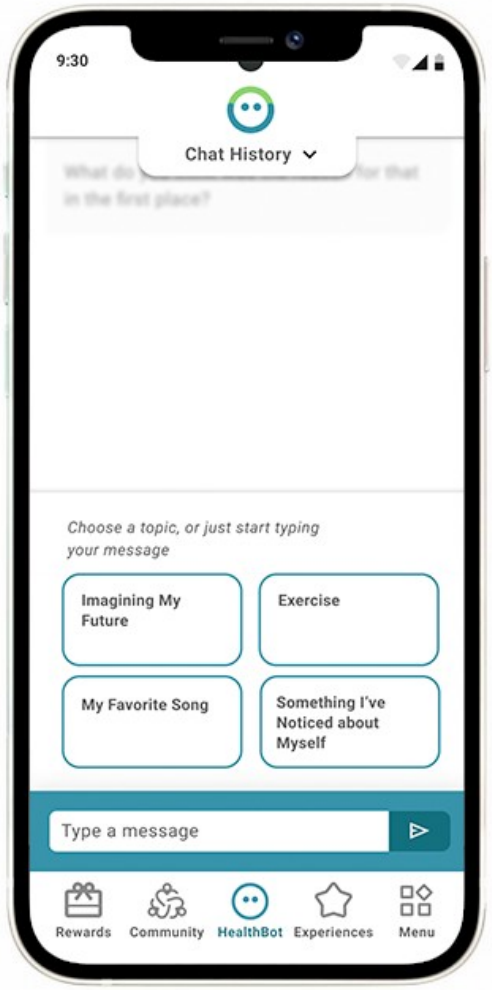
Notes + Future Steps

The unmoderated testing format limits how much we can infer about users' taps in the 'Menu" area and the LoreBot Text field.

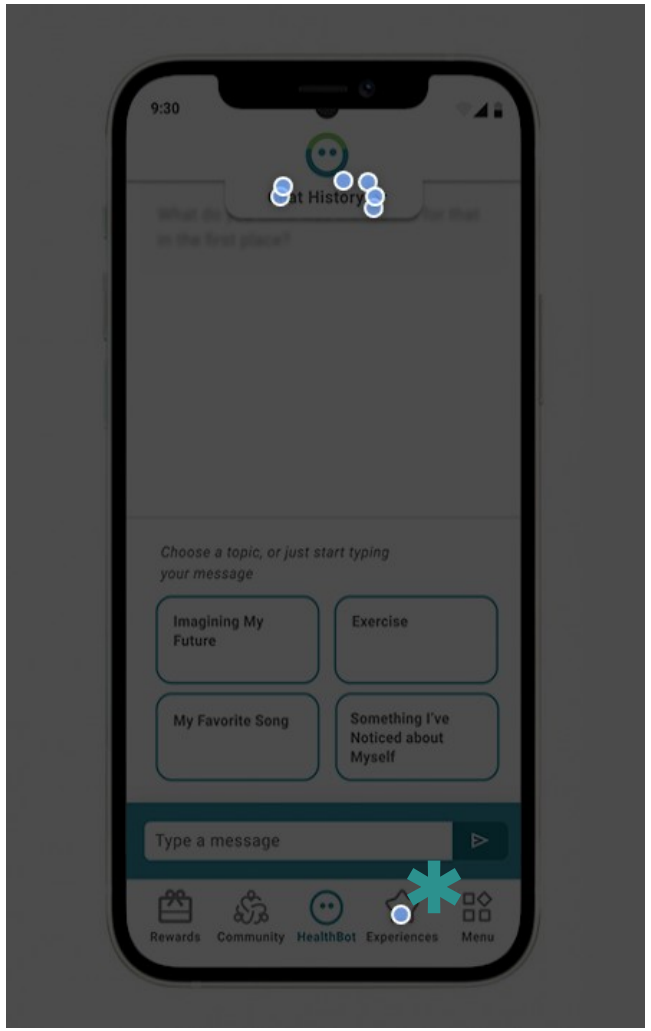
When we expand on these, we can do more isolated tests on the usability of the pills and its UI.

Triangulate this data against future tracking, and in-the-field observation

User Scenario 3
Recall a past conversation



Test Screen



Users' Tap Pattern (7 total)

Results

6 of 7 tapped "Chat History" UI

Strong User Confidence (avg. 5 / 5)

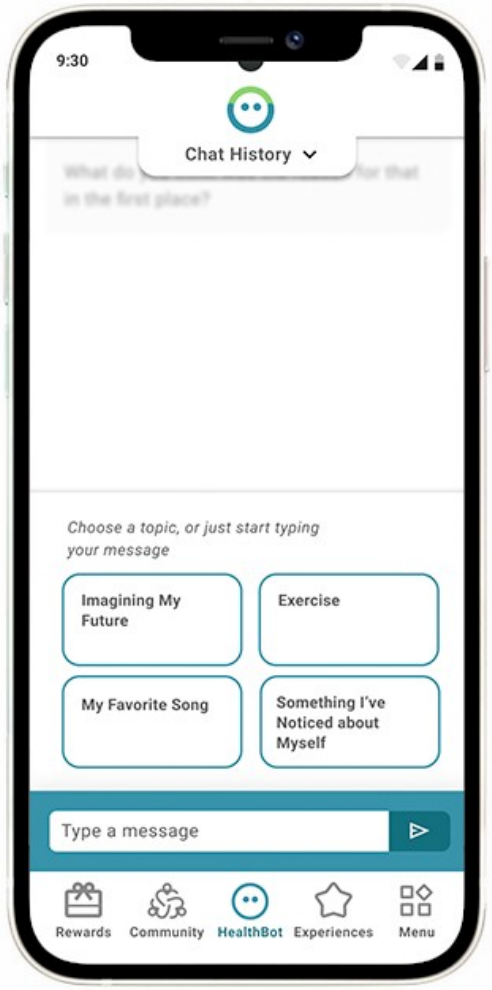
Notes + Future Steps

Measure open/close usage, as well as test how users navigate and understand the upcoming feature

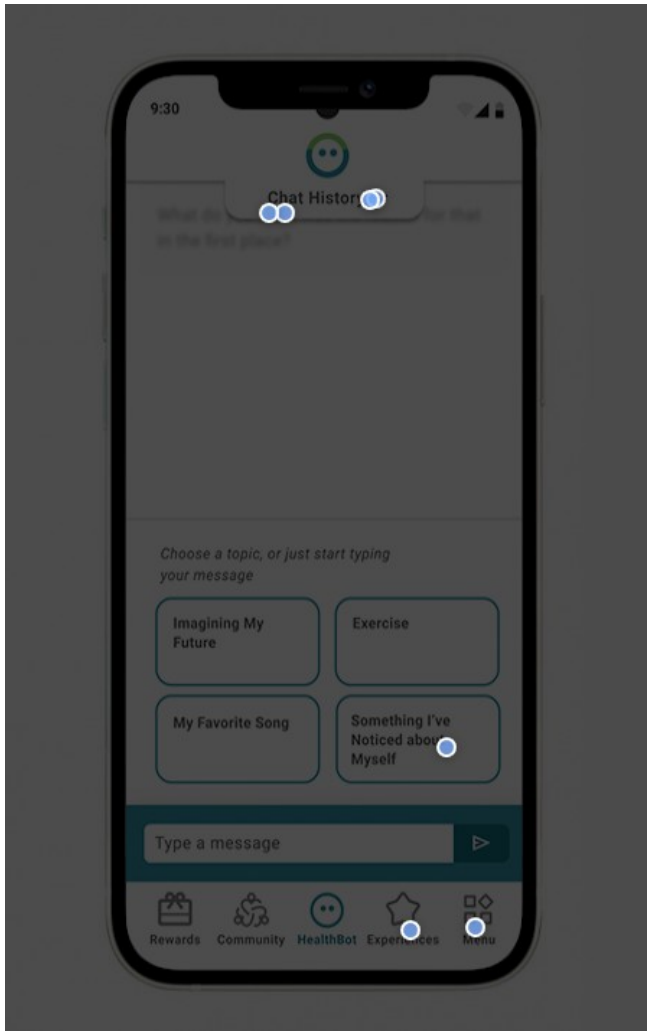
* The outlier tapped "Experiences" logically makes sense, so the user was looking AND thinking. But if the designer would have updated it to "Explorations" this would not have happened!!

Triangulate this data against future tracking, and in-the-field observations

User Scenario 4
Return to a chat from earlier today



Test Screen



Users' Tap Pattern (7 total)

Results

0 of 7 tapped "Chat Transcript" / Text

Mixed User Confidence (avg. 3 / 5)

Notes + Future Steps

This subtle version with minimal text likely is too subtle.

The current prod version shows much more/ very bold chat transcript, and might be a good set of 'bookends' to find a good visual balance.

Consider other styling like gradient, fade in, scrolling as loading to indicate action.

Triangulate this data against future tracking, and in-the-field observations

Next or Future Steps

Measure
Evaluate again when full features exist

Method
Triangulate with front-end click data and in-the-field observation

Chat Transcript
Evaluate with analytics in the current, bold form

Consider Pill Content
Trial and error/success with different topics, for those without stored data and those with personalized topics

Thank you!

Appendix

Future Tests (All)

Other Testing Needs/Ideas

Preference Testing on Pill ‘Types” (BE)

Do preference testing between generic mood-based or custom.

Ask “why” in a scenario.

This might be challenged by limit of testing/hypothetical/scenarios

We might start with data since it’s running, and no cost and see if the A/B test shows a distinction. (And we just keep it running all the time to be purposefully random.)

Example:
‘Something light’, vs from user history/past

**(BE) When users have questions about what is lore, FAQ?
Q+A format? Video?
LoreBot answers?
Chat with Eric? Shake to report or feedback in Native?**

Do preference testing around method of serving up help, or push, pull etc?

This likely wouldn’t be a usability thing, but a mechanism to serve up help, or research into when and where users are in the app, their usage journey, etc

Why We Observe

Why We Observe

We observe to understand behavior

Observation is different from discussions, interviews or surveys because:

Human memory is fallible

People and AI want to please people

When it's 'users' they are pragmatic and concrete

We also observe to

build a culture of understanding about our customers

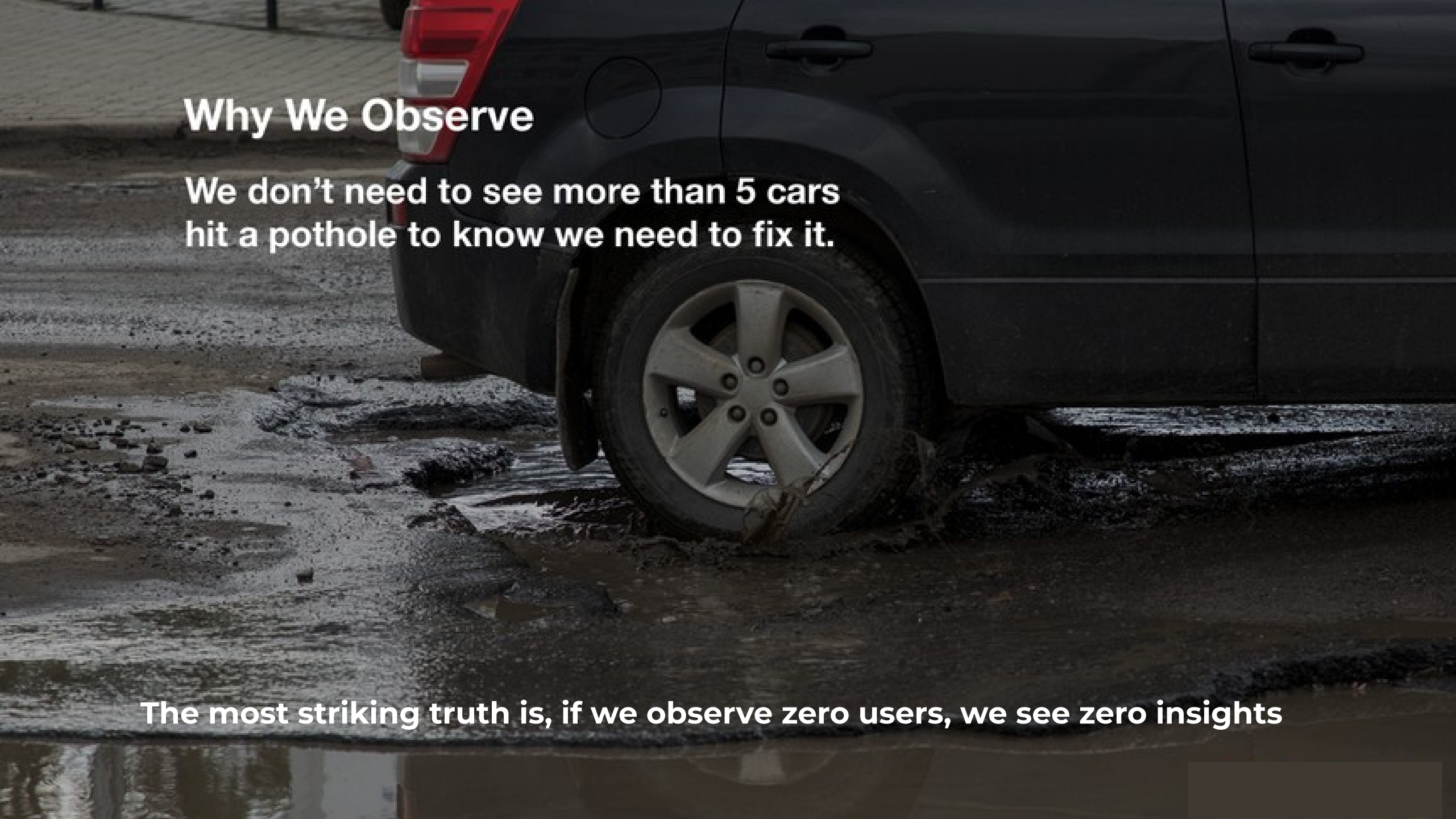
Why We Observe

85% of usability issues can be uncovered with 3–5 participants



Fix the issues, and test again

https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/



The most striking truth is, if we observe zero users, we see zero insights

What is Lyssna?

Lyssna (Usability) Tools

What usability testing is

Directional insight (are they navigating, interpreting as we intended)

Finds mismatches with our design and user mental models

Focused on functionality, clarity and ease of completing a task

Short term, great for Evaluative Research

What usability testing isn't

Statistically significant

Not a measure of
a feature value or need

Doesn't replace market research, surveys or interviews

Longer term methods like diary studies are great for Generative Research to learn new insights and ideas around benefits

Mixed methods research gives us a 360° view

Lyssna Core Testing Features

Quick Iteration Tests

First Click: where do users *instinctively* go to do start a task

Five-Second Test: what do users *recall* after a short exposure to the stimuli

Preference Test: expose users to multiple versions* of a screen, interaction, design for subjective response

Task-Based Tests

Prototype Test: *observe* users completing tasks (like Figma)

Navigation Test / Tree Testing: see how users pragmatically *find, search, navigate in a task* and inform our information architecture (menu, settings, labels)

Card Sort: understand how users *comprehend* groupings and label content

Other areas of testing outside of Lyssna

Live Site Tests

Moderated usability from our current users. This is the most efficient, data-rich testing method.

(It's a challenging path to execute right now.)

Other Tests

System Usability Score (SUS) (Popular but limited in insights)

Marketing and email campaign testing (Lyssna does have that ability.)

Market research or survey integrations (Are there 'off the shelf' options?)

Usability Testing Creates

Design Confidence

Small-sample feedback helps reveal patterns and friction early

Workflow Validation ('Validation' has a limited meaning in usability testing)

Did users complete an intended flow? If not, we adjust structure or clarify.

Iteration Fuel

Refining direction, and each round informs us. Less about right and wrong, but "better".

Squad Alignment

Everyone has a shared view of user behavior, making design decisions less subjective.

Limitations in Usability Testing

Time-based or scenario-based

Won't capture any emotional or reflective goals

Won't 'prove' an idea, or a UI over another, but we can learn how to improve.

We will have to infer some things based on how users approach / mental models. (So some post-test analysis on our part.)

Limited ability to ask questions in an asynchronous testing session (we can ask follow ups though.)

It's the difference between interface understanding and human experience

The Exponential Cost of Late-Stage Discovery

Fixing a problem in development costs 10X more than fixing it in design.

And 100x more if discovered after a release*.

Stage	Relative Cost of Fix
Design	$1
Development	$10
After Release	$100

Lightweight usability testing is just not helpful, it's financially strategic.

** Bruce Togazzini, founding member of Apple's Human Interface Group, "First Principles of Interaction Design"*

########

Shared Test Results Link: (Lyssna)

Date	Duration (ms)	Age	Country	Gender	Education level	Annual household income	Employment status	Daily hours online	Technical proficiency	Device Type	Device Platform	Device Width	Device Height	1. (First click test: "You already know what you want to talk about, and yo
2025-06-06 15:24:04	111431	77	US	Male	Some college	$40,001–$70,000 / year	Part-time employed	1–3 hours	Advanced	mobile	ios	390	656	14011
2025-06-06 15:29:19	66979	35	US	Male	Post graduate degree	Over $100,000 / year	Full-time employed	Over 6 hours	Advanced	mobile	android	360	680	7230
2025-06-06 15:30:27	54724	25	US	Female	College graduate	I'd prefer not to say	Full-time employed	4–6 hours	Advanced	desktop	mac	393	666	3105
2025-06-06 15:30:51	95346	56	US	Female	College graduate	Over $100,000 / year	Full-time employed	4–6 hours	Advanced	mobile	ios	390	665	7262
2025-06-06 15:34:37	112637	57	US	Female	Post graduate degree	Over $100,000 / year	Full-time employed	Over 6 hours	Advanced	mobile	ios	393	647	9480
2025-06-06 15:45:00	80550	53	US	Female	College graduate	$40,001–$70,000 / year	Self-employed	Over 6 hours	Advanced	mobile	android	384	731	7097
2025-06-06 16:03:53	174906	30	US	Female	Post graduate degree	Under $10,000 / year	Unemployed, seeking employment	4–6 hours	Advanced	mobile	ios	402	685	12410

