

Rate your Lyft

Weiran Ma, Marlo Owczarzak, Yun-Han, Mingchu Cong



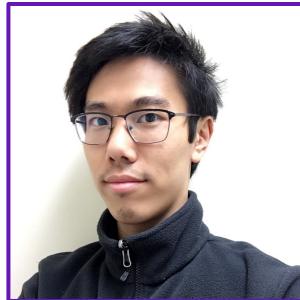
Table of contents

4	Introduction	36	Initial Wireframes
5	Solution Summary	44	Usability Testing
6	Secondary Research	48	Iteration
20	Primary Research	49	Final Design
28	Personas	56	Lyft Rating Video
30	Problem Identification	57	Reflection
32	Design Goals	58	Conclusion
33	Ideation	60	References & Appendices

Team



Marlo Owczarzak



Weiran Ma



Yun-Han Huang

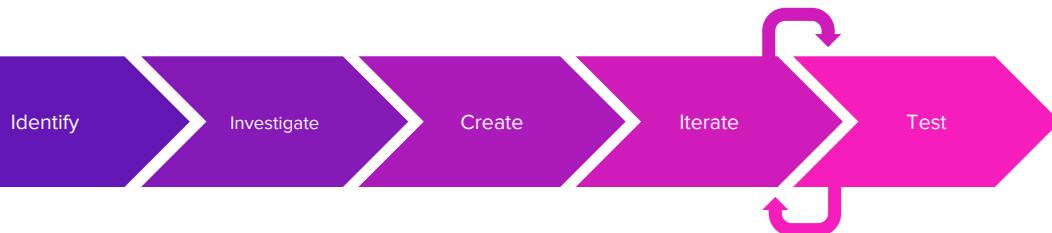


Mingchu Cong

Introduction

Background

The “sharing economy” is one of the fastest growing concepts all over the world (Cohen & Kietzmann, 2014). Lyft, the leading ride-sharing provider, is now serving thousands of people in the U.S. every day (Bensinger, 2017). However, due to the limitation of its five-star rating system, the current ratings of lyft drivers may not be truly reflecting rider’s experiences, as well as the overall quality of the services provided by drivers. Thus, the goal of the project is to design a new rating system which can effectively match the quality of the service provided by drivers with the experiences perceived by the riders during their trips.



Process

In this document, we will first present how Lyft’s current system works, and how its 5-star systems would limit users’ judgement of the quality of the services provided to them through primary and secondary research. Second, we will discuss how our new system would change users’ perception of their riding experiences hence encourage them to make correct rating decisions. Finally, we will present issues found in our usability testing as well as refined solutions based on feedback provided from our participants.

Solution Summary

Our final solution adopts a binary rating system, which records the positive and negative feedback through users' ratings with thumbs ups and thumbs downs at the end of each trip. If users would like to leave neutral feedback, they could also submit payment for their trip without leaving any feedback. Additionally, once users provide their general impressions of their trip, they are encouraged to select the reason(s) why they gave such ratings, which would help drivers better understand what aspects they did well and what aspects they should improve. It also prevents malicious users to give a negative overall rating without providing any detailed feedback.

The wireframe shows a user profile for "Tom Neilson" with a placeholder icon. Below the profile are buttons for "No Tip", "\$1", "\$2", "\$5", and "Other". A "Payment" button is shown with a credit card icon and the text "VISA **3655". The total amount is displayed as "\$11.90 credit applied" and "\$12.90". Below the amount is a "Rate your experience." section with "Like" and "Dislike" icons. At the bottom is a large "Submit" button.

This wireframe extends the previous screen. It includes a "Rate your experience." section with "Like" and "Dislike" icons. Below this is a "Give us some detailed feedback. (Optional)" section with four buttons: "Safety" (green), "Navigation", "Friendliness", and "Cleanliness". At the bottom is a text input field with the placeholder "leave comments here ...". A large "Submit" button is located at the very bottom right.

Secondary Research

Demographics

Pew Research Center provides demographics about American adults using ride-hailing applications in 2015. As the graph shows, 15% of American adults have experience using ride-hailing applications. 51% of American adults heard of the services but have no actual experience. 33% of American adults never heard of these services.

Among those people who use ride-hailing applications, ride-hailing is a relatively sporadic activity. 26% of the users use these services on a monthly basis, and 56% use them less often monthly. However, 17% users utilize these services more frequently: 3% use on a daily or near-daily basis, 14% use weekly.

The statistics drawn from the above research informed us that the majority of American adults have not actually done any ratings on ride-sharing apps, which means that they are lack of understanding of how the rating system works and may not have an agreement on the standards of defining a good and bad driver.

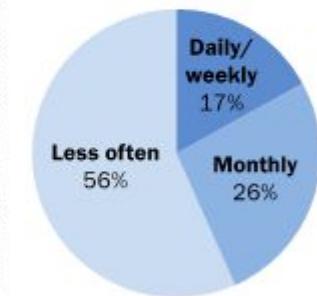
Picture: Demographics about American adults using ride-hailing apps
Retrieved from: <http://www.pewinternet.org/2016/05/19/on-demand-ride-hailing-apps/>

15% of American adults have used ride-hailing apps

% of adults who ...
ride-hailing apps



Among those who use ride-hailing, % who use them ...



Source: Survey conducted Nov. 24-Dec. 21, 2015.
"Shared, Collaborative and On Demand: The New Digital Economy"

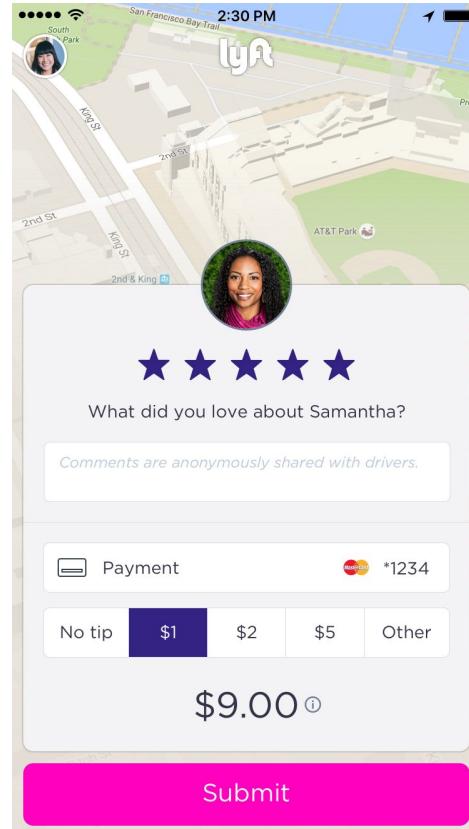
PEW RESEARCH CENTER

Secondary Research

How does Lyft rating system work - mechanism

Lyft currently uses a five-star rating mechanism. The driver rating is the average of the driver's last 100 ratings, or however many the driver has received so far.

Further detailed feedback about safety, navigation, friendliness, cleanliness and typing in comments are optional for the passengers.



Picture: Lyft rating UI

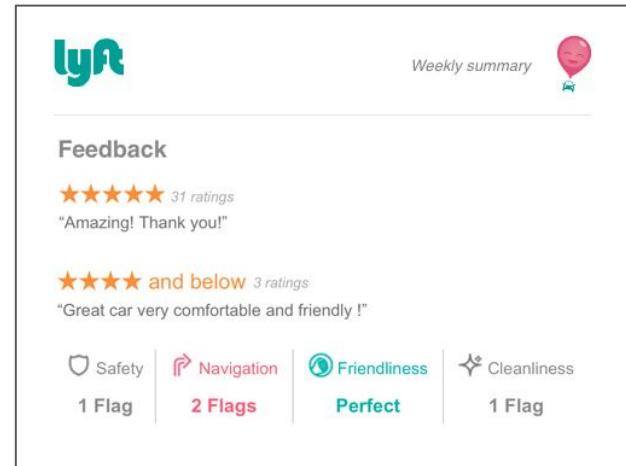
Retrieved from: <https://help.lyft.com/hc/en-us/articles/214582857-How-to-rate-your-driver>

Secondary Research

How does Lyft rating system work - driver report

A driver receives weekly reports about passenger feedback and flags. This report is meant to be a guide helping the driver on improving.

Flags include: navigation (the most common), safety, cleanliness, and friendliness.



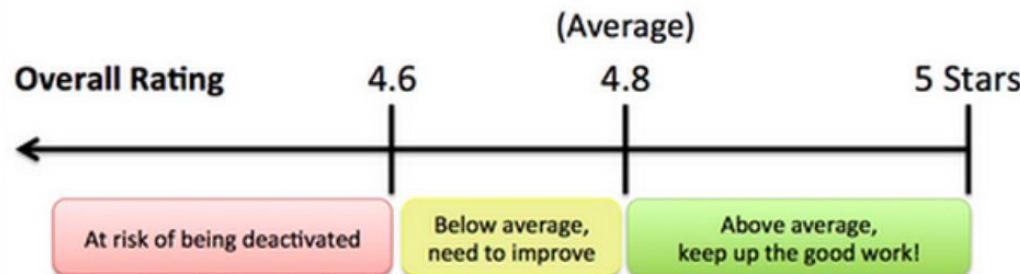
Picture: Lyft summary report for a driver

Retrieved from: <http://www.kellydessaint.com/tag/lyft/>

Secondary Research

How does Lyft rating system work - interpretation of ratings

According to Lyft's official description of a "Good Driver Rating", anything more than 4.8 is awesome. If the rating drops below 4.8, the driver may want to consider ways to improve it. Consistently low ratings (below 4.6) can put the driver at risk of deactivation.



Picture: Lyft rating standard

Retrieved from: <http://static1.businessinsider.com/image/54db8911dd08959a1f8b45a4-952-300/screen%20shot%202015-02-10%20at%202016.36.04.png>

Secondary Research

Competitive analysis

	Lyft (new)	Uber	Lyft (old)
rating system	binary	5-star	5-star
feedback	Comments + Tags	Comments + Tags	Comments
tipping	available	unavailable	available

Secondary Research

Cyber trust

With the occurrence of e-commerce and sharing economy, people take risks when conducting transactions with strangers. People's willingness to take this kind of risk depends on their trust in the platforms. In other words, trust is also an effective way to advance online transactions. Etzioni (2017) pointed out some factors affecting how people trust online platforms, such as familiarity and business ethics.

In Uber and Lyft, both drivers and riders exposed themselves to risk when riding with strangers. To promote trust, Uber and Lyft established their rating systems. **However, Cook (2015) showed that among the ratings in Uber, only 5% are three-stars or lower. Etzioni (2017) posed an issue that both Uber and Lyft's ratings are inflated.** Usually in a five-star rating mechanism, 2.5 is average and 5 means outstanding. However, from Uber's data it shows that five could mean anything from average to outstanding.

Secondary Research

The purpose of rating

Online reputation systems, also known as rating systems and reviewing systems, are very common in people's online activities nowadays. According to Rainie, L. et al. (2004), these kinds of systems enable people in making decisions on who to trust, comparing opinions with others, and elevating accountability of an institution.

In a report written by Anderson, M. (2014), it showed that 88% of consumers trust online reviews as much as personal recommendations. **The number shows people's attitude toward online reviews and the importance for Lyft to establish a reliable rating system in order to maintain the company's reputation.**

Secondary Research

Polarized Ratings

Allen (2006) stated that Undetailed Rating (where the rater isn't required to add additional information other than the rating they select) would lead to a bimodal distribution of ratings trends, meaning that users would either give a 5 or 1 star, and avoid giving in-between ratings. However, Lyft's commenting feature does not change the fact that passengers still tend to give extreme ratings on a 5-star system. Hu, Pavlou, & Zhang (2006; 2009) further argued that the cause for this behavior is because people are less motivated to express (or complete a rating) if they have neutral views over a product or service (*under-reporting bias*), while those who are extremely satisfied or dissatisfied are more motivated to express their views (*purchasing-bias*).

Thus, the results from the above research indicate that scores on a holistic five-star system do not accurately reflect the entire user populations' true opinions towards the product or service. This affords us to think of possibilities to create a rating system that would effectively avoid users' polarized rating behaviors, and truly reflect the quality of riders' experiences.

Secondary Research

Rating systems

Four different rating scales:

unary (“like it”), binary (thumbs up / thumbs down), five-star, and a 100-point slider.



- Users' rating costs increase as they have more rating choices. All scales show similar cognitive load. However, page rating times increase significantly with finer-grained scales - despite users leaving more items unrated with those scales.
- User rating times between the unary and slider vary more in relative time for movies (62% vs 25%), but vary more in absolute time for reviews (about 2 vs 3.5 seconds).
- Users spend more time assigning ratings at the middle of a scale, such as three or four stars on a five-star scale.
- Users prefer the five-star scale overall, although the thumbs scales comes in as a relatively close second choice for product reviews.

Secondary Research

Current problems in Lyft rating systems

In many blog posts, online forums, and news, there are lots of drivers discussing suffering from the current rating system, and complaining about the rating issues:

- Different riders have different standards on five-stars rating
- Picky riders affecting the rating / riders not understanding how rating works
- The mechanism being too harsh to the drivers
- The drivers have no idea how to reach high ratings

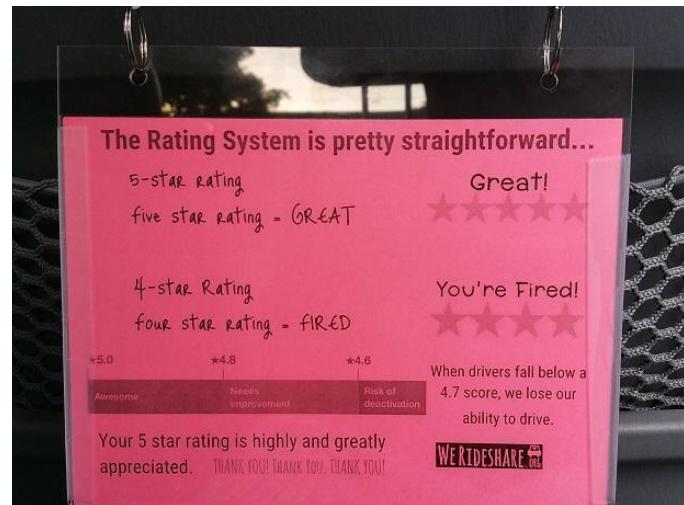
Secondary Research

Current problems in Lyft rating systems

Problem 1 - Different standards

“The feedback system is gravely flawed as it has no pre-defined standards.”

“I would say, ‘Hey I’ve got a question. You know the ratings system, do you think 4 stars is a good rating?’ About two thirds said, ‘Yes.’”



Secondary Research

Current problems in Lyft rating systems

Problem 2 - Picky riders affecting the rating / riders not understanding how rating works

"There are always passengers who are nitpickers who will rate the drivers based on the most trivial, insignificant and often illogical things."

"Often if a passenger is new to the platform, they will assume that the rating is about the company, not the driver."

"Let's say you're a 4.8 rated driver and one passenger is in a really bad mood. Let's say you did something minor, like maybe took a wrong turn, or made some offhand comment that was taken the wrong way, or you didn't notice a mess a prior passenger left in the back seat. Even though you didn't do anything really bad, they still rate you one star."

Secondary Research

Current problems in Lyft rating systems

Problem 3 - The mechanism being too harsh to the drivers

"So let's see... to recover from one 1 star rating, you need 19 five star ratings. Yes, that is a very harsh system."

"In my poll, when I told passengers that a driver with 4.6 is likely to be de-activated, they were amazed. One passenger said, 'Oh I feel so bad now. I almost always give drivers 4 stars – unless they do something really amazing.' "

The screenshot shows a post on the subreddit r/Lyft. The title of the post is "New Driver Struggling with Ratings" and it was submitted 2 years ago by the user Chagost. The post content reads: "I've been driving for about 3 weeks now and have completed about 35 rides. I fired the first time this morning and it was pretty low (4.5) and I was wondering if anyone has any advice. Unfortunately, I have seen no specific complaints so I don't know what I am doing to have appreciated." Below the post, there are upvote and downvote arrows, a comment icon, and a link to 19 comments.

"Yes, ratings do put a kind of fear in the drivers and keeps them in line. That's the intention, but is that fear factor really necessary as a motivational tool? Do we have any examples of systems that don't use ratings? What would happen without ratings? Would Uber's level of service suddenly devolve, a downward spiral ultimately bottoming out somewhere near taxis?"

Secondary Research

Current problems in Lyft rating systems

Problem 4 - Have no idea how to reach high ratings

"Providing a high quality of service to riders, what does that even mean in concrete terms? Where is the official guideline? Do they plan on creating a standardized instructions for drivers and passengers?"

I am so pissed about the Lyft rating system

Discussion in 'Philadelphia' started by OSC, Mar 31, 2017.

Page 1 of 2 1 2 Next >



OSC
Active Member
Location: Philadelphia
Driving: Lyft

1-My car, a new 2015 SUV, leather seats.
2-Inside out all clean, very clean actually.
3-Spray high end perfume inside (not too much, just a little bit to make it classy) like Chanel alike.
4-Pick up on time, or faster then expected, right where they live/stand. Drop right in front of their door/destination.
There was no issue for my yesterday's work, but my rating has dropped from 4.8 to 4.7.

What the hell should I do to to get 5 star? lick their feet or what?

About me: clean, wearing formal clothes when driving. Because I do Lyft right after my day job so I don't go home and change to jeans and shirt. Great and smile everyone of them when they enter and leave my vehicle, talk to them if they like to talk.

About 9,240 results FILTER


5 Tips to ALWAYS Get a 5 Star Rating
Marc Frecero • 12K views • 8 months ago
A big question I get regarding driving for Uber & Lyft is techniques to get a 5 star rating from passengers. A high rating ensures ...
4K


New Lyft Rating System And Phone Support Coming Soon!
Your Driver Mike • 1.6K views • 2 months ago
Lyft just announced a new rating system and improved phone support coming soon! How can Lyft (and Uber) improve their current ...
4K


How to improve from a 4.5 uber and lyft rating
Josh Bar • 12K views • 1 year ago
If you find this helpful throw me a bone and use my referral code. <https://www.lyft.com/drivers/JOSHUA4487> The rideshare driver ...
18:09


Why Uber/Lyft's Rating System Sucks.
thinkFAST • 2.2K views • 1 year ago
All the reasons why the Uber/Lyft rating system is pretty much the worst thing ever. We're on social media: ...
2:52

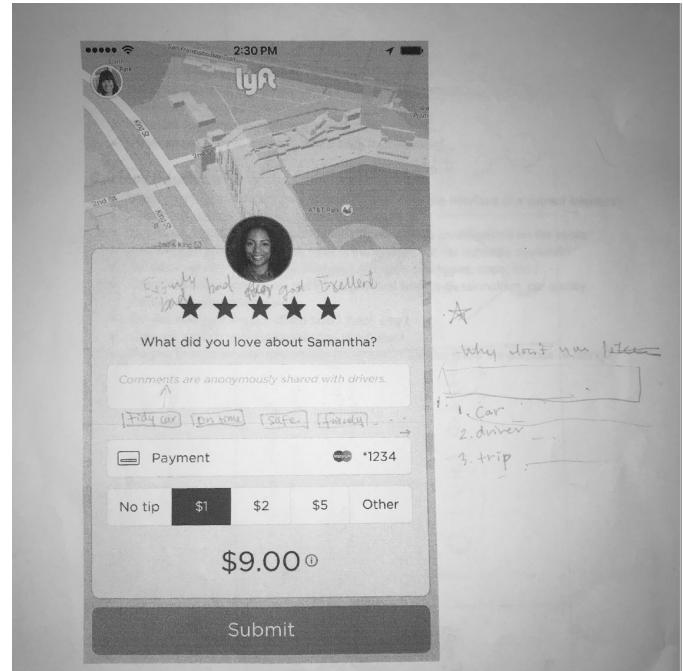

#1 Tip for Consistent 5 Star Ratings for Uber or Lyft!
Drive Girl Drive • 38K views • 11 months ago
I have maintained a 4.94 rating for 10 months. This is my best advice on how to maintain and keep your ratings as high as ...
10:03

Primary Research

Method - Interview & design workshop

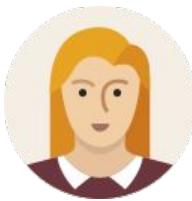
We recruited five participants for **semi-structured** interviews and a **design workshop**. We audio recorded the interviews and converted these recordings into transcripts for further analysis.

We also conducted a design workshop with our participants. Hannington & Martin (2012) stated that design workshops are efficient, compelling, fun ways to gain the creative trust and input of stakeholders through activity-based research. Thus, we provided some printout of current Lyft user interface for the participants to present their ideas on the papers.



Primary Research

Participants



Female

28



Male

24



Male

21



Male

26



Male

24

Primary Research

Interview objectives

The purpose of interview is to understand participants':

- Overall impressions of Lyft
- User behavior on the Lyft application
- Previous experiences on taking Lyft rides
- Perceptions on the current Lyft rating mechanism
- Experiences and impressions about other rating systems

* Please see Appendix A and B for our interview protocol and questions

Primary Research

Affinity diagramming

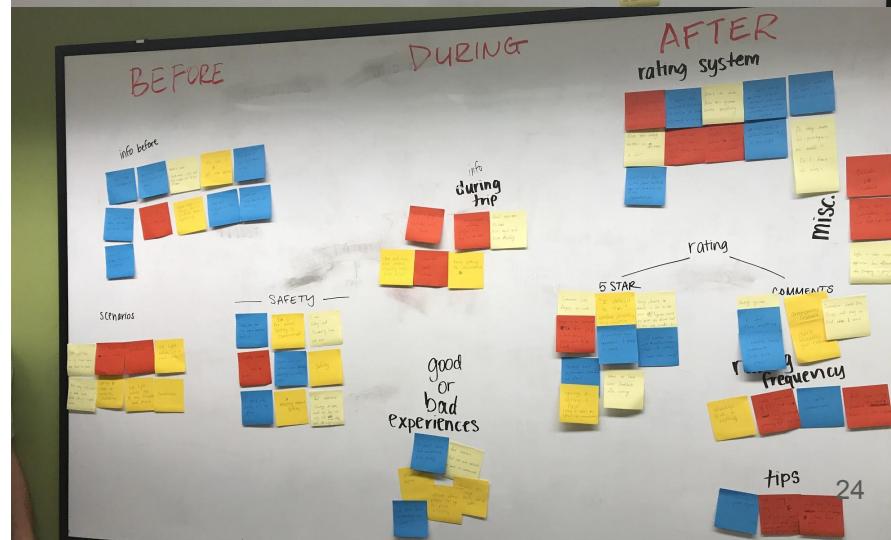
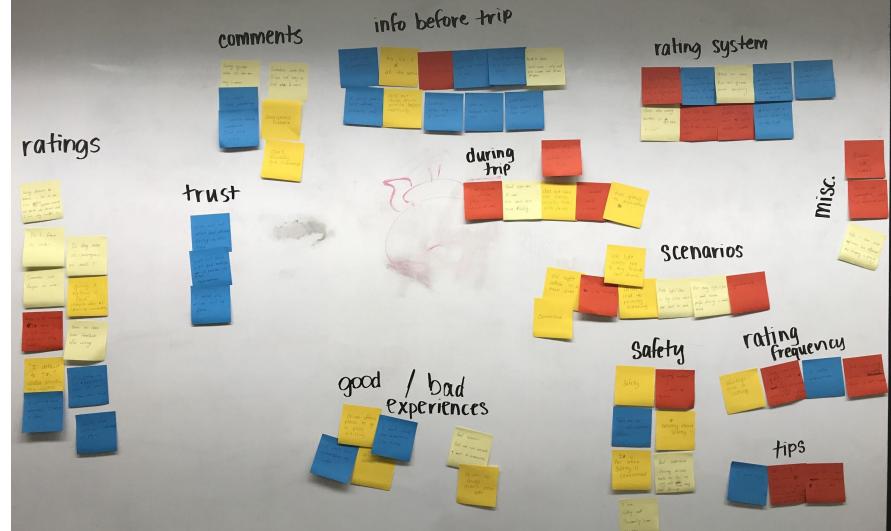
We organized information from our interviews in an affinity map. The map transformed over several iterations as we combined and restructured sticky-notes. The top image has the information organized by interviewee, and the lower image shows the information in our first attempt at categorization.



Primary Research

Affinity diagramming

The top image shows the first attempt at organizing information, and the lower image shows a later iteration. In addition to finding similarities in concerns, we also found that our interviewees had different concerns before, during, and after a Lyft ride. These distinctions were clear, so we further organized our information into three columns.



Primary Research

Highlighted Interview Quotes & Insights



"I don't normally rate, unless it is a really good or bad experience."



Insights: users avoid giving feedback to mediocre experiences



"Ratings above 4.5 are all the same to me, there are always people who will rate badly for a bad reason that they found irritating, but isn't pertinent to the safety or destination."



Insights: users don't see a difference between star ratings



"I don't check the driver's rating. Lyft is supposed to deal with the issue. Or it will harm the company's reputation."



Insights: users trust Lyft and Lyft drivers' quality, they don't really care about the driver's rating



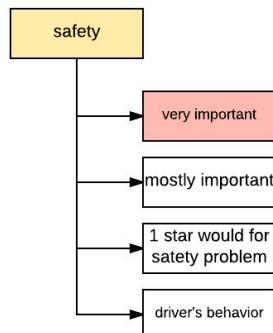
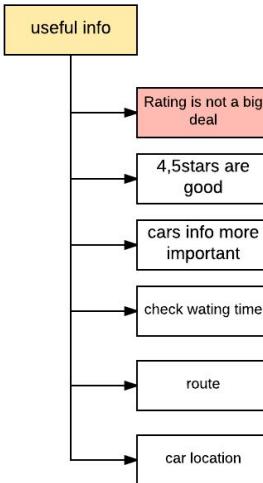
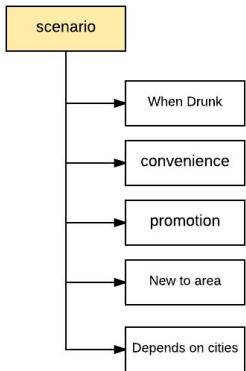
"I don't actually know. I think the system just average of the week. I don't have a concrete idea."



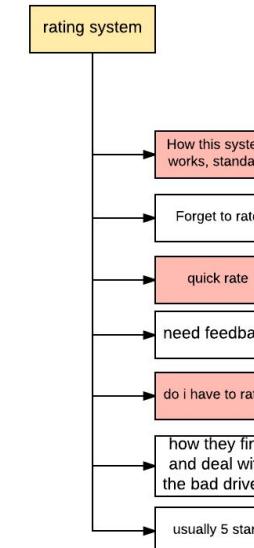
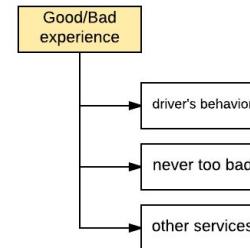
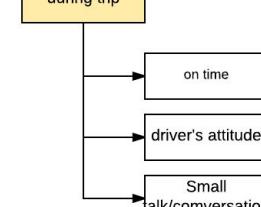
Insights: users do not have a concrete concept about how Lyft calculates ratings and deactivate drivers.

Primary Research

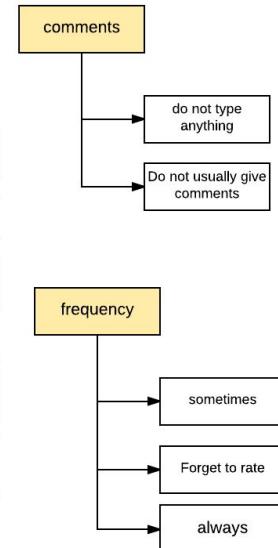
Before



During



After



Primary Research

Summary of Insights from interviews

- Impression on Lyft / Trust in Lyft
 - Most participants have good experiences in rides and have positive impressions about Lyft.
 - Their main concern is mostly about safety, but they think Lyft is reliable.
 - They believe that Lyft inspects on the drivers and doesn't allow dangerous people driving Lyft.
- Rating behavior / knowledge / perception
 - Participants are familiar with rating systems such as Amazon and Yelp. Rating is very common to them.
 - Ratings of 4.6, 4.8, 5 do not mean differently, they all means good drivers.
 - With the exception of one participant who knows that drivers below 4.6 might be kicked out, most participants are not really sure about how the mechanism works.
 - Most of the participants rate five stars all the time.
 - Most of the participants don't check a driver's rating carefully, and some even don't really care.
 - Some participants rate every time, some do not. Some participants especially hate typing in comments.

Persona

Basic info

Soo-Jung is an international student from Korea, he is a first-year graduate student majoring in Engineering Technology, and he is quite satisfied with his experience here so far. He usually takes advantage of holidays and breaks to travel to different places in the U.S. because he would like to go back to Korea for employment since all his family members are still in Korea. He owns a car, but he never takes on a road trip using his own vehicle. He would normally fly to the city he visits and rent a car from three. Occasionally, he would use lyft to get around when he travels to larger cities where car rental costs and parking fees are somewhat high. He is a bit shy and does not speak perfect English. Thus, when he takes a Lyft, he would normally sit in the back and prefer not to initiate a conversation with the driver unless he has questions to ask the driver.

Behavior

- be active on the Internet
- can't live without mobile phone
- act on emotion

Soo-Jung Park



"I would normally just sit quietly in the back and relax, thinking of things I have to do next."

Age: 26

Occupation:
graduate student

Introverted

Hardworking
Quiet

Scenario

Soo-jung would like to visit Los Angeles during his winter-break to run away from the coldness in Indiana. After he arrived at the Airport, he decided to take a Lyft to the hotel because taking Lyft is cheaper than taking a taxi. He does not want to take a bus or subway either even though they are cheaper than taking a Lyft because he felt like he did not know the city well and it was late, and all the buses and subway services were about to be shut down. He opened the app on his phone and found there were several Lyfts available around him. He set his destination and chose the cheapest option and set up a pickup. During his wait time, he noticed that the driver has a rating of 4.6, but he did not think too much of it because he does not know what rating other drivers have. But he felt like it was a good rating based on his standard as a student because it was only 0.4 points away from perfection. After the driver arrived and picked him up, he confirmed with the driver regarding the destination. He noticed that the estimated arrival time is about an hour ahead, so he decided to close his eyes to take a rest in the back seat. When he arrived at his hotel, he thanked his driver with kind words, but did not leave a tip since he noticed that Lyft, unlike taxis, does not require passengers to leave a tip, and tipping is not a tradition in his home country. He ended the trip by submitting with a 5-star rating and no-tip option. Normally, he would leave a 5-star rating to most drivers. If he is not completely satisfied with his trip, he would just force quit the app, and do not rate the driver. He only rated a one-star to a driver that he thinks is really rude and deliberately prolonged the journey because he is new to the city. In addition to the 1-star rating, he left a very long comment and to express his anger and dissatisfaction. He finally received a \$30 coupon and a reimbursement of that trip as well as a sincere letter of apology from the Lyft customer service.

Persona

Basic info

Molly is a college student enjoying traveling a lot. While traveling, her first concern is safety. She has lots of friends using Lyft. They recommended Molly Lyft since they had great riding experiences. Since that, Molly takes Lyft when she visits a new place to move around. Molly is full of curiosity and outgoing. She enjoys chitchatting with Lyft drivers to make friends and know more about the place she is visiting.

Molly is a heavy user of mobile phone and social media. She chats with her friends, and she is active on many online communities. She loves expressing her ideas by typing and interacting with people on the Internet. As the rating system of Lyft, Molly doesn't take it seriously. Also, she has no idea how the rating mechanism works and how her ratings affect the drivers. Sometimes she does the rating depending on her mood. At most cases, she rates five stars since she had a nice talk with the driver. If she really enjoys the talk or she is in a great mood, she spontaneously leaves some positive comments to the drivers. However, sometimes she rates the drivers low irrationally, for example, when she has a bad day or the driver debates with her. She feels like rating is a way to release her emotions and reflect her feelings. Even so, she never rates below three, she thinks it is too rude and she haven't had such a bad experience which worth a rating below three.

Molly Hunter



"I care about safety and I trust Lyft since it inspects the drivers and tracks my route. The best thing is that the drivers are always kind and supportive."

Age: 22

Occupation:

College student

Outgoing

Sociable

Communicative

Behavior

- be active on the Internet
- can't live without mobile phone
- act on emotion

Scenario

When traveling, if Molly needs to move from place to place, the first thing comes to her mind is Lyft. She chooses the basic option instead of Plus, Premier, and Lux since she usually travels alone or with a small group and this option is always the cheapest one. Molly opens the Lyft app in her mobile phone and allows Lyft to determine her location as the pick-up place. Once paired with a driver, she takes a glance at the driver's profile picture, name, and rating. If the picture doesn't look weird and the rating is above 4, she confirms the ride. During the ride, Molly enjoys seating in the front passenger seat and talks nonstop with the driver. As long as the driver takes her to the destination safely within a reasonable time, she doesn't really notice and care the route and driving behavior. Molly always asks the driver for anything special in the city and casually shares each other's life experience. She thinks being able to talk with the locals during the rides is a benefit of taking Lyft. The drivers always provide her helpful information and some useful tips. Once Molly reaches the destination, she takes 30 seconds to one minute to rate the ride. If nothing goes wrong, she doesn't look back the experience carefully and quickly rates a five star then exits the application. If she has a great talk or is in high spirit, she paid extra tips and leaves some positive comments. However, in some cases she is in a bad mood or she personally doesn't like the driver during the talk, she rates a four star. But she doesn't explain much on that.

Problem identification

- **Different standards**
People have different understandings of one to five stars.
- **Rating mechanism**
People have no idea about how the rating system works, how their ratings affect the driver (they don't know drivers below 4.6 will be kicked out).
- **Rating inflation**
Usually in a five-stars rating mechanism, 2.5 is average and 5 means outstanding. However, in Lyft, 5 can mean anything from average to outstanding, which is conflicted with people's common sense on five-stars rating.
- **Time consuming**
Leaving comments is time consuming in the context of taking Lyft. For example, passengers rush to an event or carry lots of stuff in their hands.
- **Data of rating system is not very helpful for other users**
Different people care about different aspects of the ride when taking Lyft. An overall average number doesn't provide useful information. Also, many people don't care about the ratings at all.

Problem statement

The current Lyft five-star rating system is inefficient and ambiguous to the riders.

The current system does not promote concise feedback to the driver to improve their ratings.

Design goals

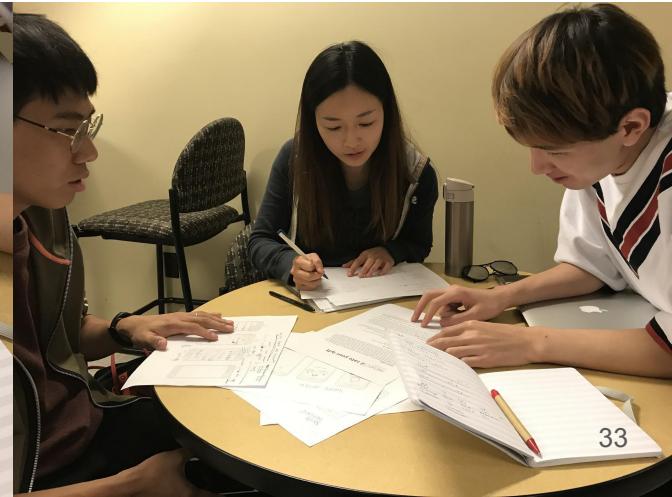
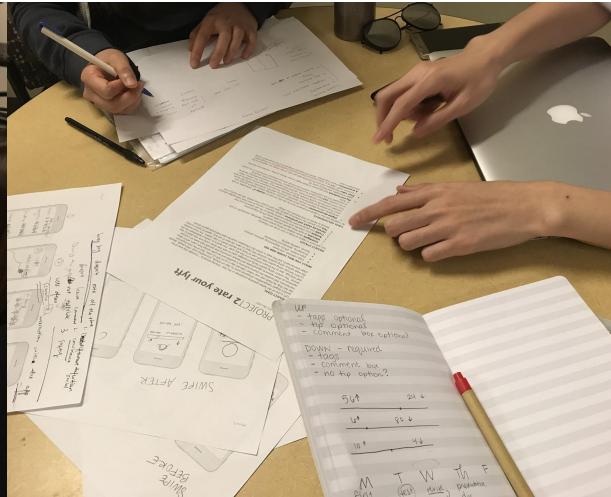
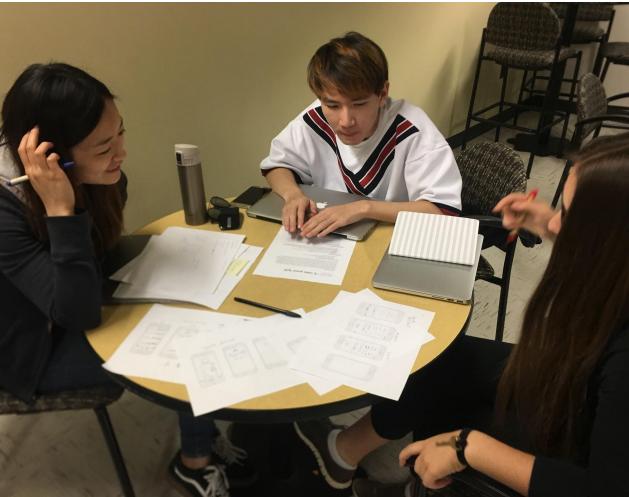
Based on the problems we identified, we set up the design goals for new Lyft rating system:

- **Efficient**
The rating process should not take the riders too much time and effort, but should reflect their feedback outlining the reality of the trip.
- **Flexible**
All of the ratings options are within one tap, but optional at the same time. The riders won't get stuck when they are hesitating on making decisions.
- **Descriptive and clear**
The riders should not be confused with the meanings of different ratings. Meanwhile, they have a mindset on how the rating mechanism works.
- **Trustworthy**
The rating system helps establish trust in Lyft and provide better future riding experiences.

Ideation

Brainstorming

During brainstorming, we sketched as many ideas as possible on papers and discussed the pros and cons of each idea.



Ideation

Brainstorming - design ideas

1. Swiping

Like the swiping gestures in Tinder, the users simply swipe left and right for thumbs down and up

2. Commenting Tags

Using crowdsourced tags for users to tap to reduce riders' efforts to type words into comment box

3. Descriptive rating scale

Using descriptive words such as satisfied, needs improvement, and not satisfied to replace numerical rating scale

4. Emoji rating

Similar to descriptive words, using emojis to present different rating scales. We discovered it would be hard to calculate the overall rating



Ideation

Design concepts

- **Replacing five-star rating with thumbs up / thumbs down**

If most of the drivers are getting five stars, five-star ratings are not very useful for the Lyft context. We believe thumbs up / thumbs down will be more efficient.

- **Fast and simple**

From previous research, we found that people are less willing to take the effort to rate on short-term experiences, in contrast to long-term experiences such as Airbnb. Thus, we will design a rating system that is fast and simple.

- **Flexibility of rating**

All of the rating inputs are optional. The user is free to skip any steps. We have considered if it will reduce ratings in Lyft. But for the overall rating, the user takes the same effort (one tap) for thumbs up/down and submit (neutral), it doesn't take more effort for thumbs up/down.

- **Simplified driver profile overview**

From previous research, riders don't care about driver's profile that much. We simplified the profile to the driver's name, car info, and thumbs up/down overview.

Wireframe A

Our first design included swiping left or right to indicate leaving an overall impression, and then giving detailed feedback.

The user also has the option to just press submit and give no rating.

The wireframe illustrates a three-step process:

- First screen (Left):** A survey page titled "Sorry!" asking for feedback. It includes four rating sections: Safety, Navigation, Friendliness, and Cleanliness, each with thumbs up and thumbs down icons. Below these is a text input field labeled "leave comments here ...". At the bottom is a large "Submit" button.
- Second screen (Center):** A payment confirmation screen for a driver named Tom Neilson. It shows a profile icon, the driver's name, and a tip amount of \$12.90. The tip amount is noted as "\$11.90 credit applied". Below the amount are three thumbs up and thumbs down icons, with the middle one being greyed out. A "Payment" button is at the bottom.
- Third screen (Right):** A summary or feedback page with a title "Tell me more about your experience!". It lists the same four rating categories as the first screen, each with thumbs up and thumbs down icons. Below the ratings is a text input field labeled "leave comments here ...". At the bottom is a large "Submit" button.

Below the screens, arrows indicate the flow: "Swipe left: thumbs down" points to the first screen, "First screen" is centered, and "Swipe right: thumbs up" points to the third screen.

Wireframe A

Each design was informed by our primary and secondary research. For this particular design, we decided to have the user give an overall rating of thumbs up or thumbs down. Secondary research shows that **a binary rating system is less of a cognitive load on the user**, and since this experience should be fast and convenient for the user, we moved from a five star rating system to binary. A central problem found in primary research is that users **have different understandings of what the scale of five stars mean**. We are attempting to mitigate that issue by providing the user with only two choices for overall impression. In interviews, users were confused by the meaning of a five star rating, so we also provided the option to give a binary rating for four individual categories. The categories Safety, Navigation, Friendliness, and Cleanliness already exist on the Lyft platform, but are only shown to the user if they rate a low star and want to “flag” a driver in one of the categories. By showing a breakdown of categories, we are addressing the issue of the passenger being unaware of the criteria for overall rating.

Wireframe B

Our second design displayed detailed feedback ratings and the overall rating **in one screen**. The user would press thumbs up or thumbs down for each of the four categories, and then press thumbs down, neutral, or thumbs up at the bottom to give an overall rating.

The user also has the option to just press submit and give no rating.

The difference between Wireframe A is the user see detailed feedback ratings before the overall rating.

Tom Neilson

No Tip \$1 \$2 \$5 Other

Payment

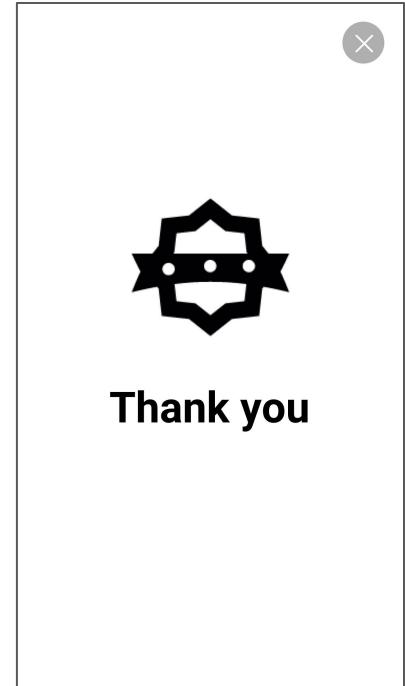
\$11.90 credit applied
\$12.90

Give us some detailed feedback(optional)

Safety	👎	👍
Navigation	👎	👍
Friendliness	👎	👍
Cleanliness	👎	👍

leave comments here ...

👎 submit 👍



Wireframe B

Through our primary and secondary research we know that it is important to keep the rating process **fast and simple**. Some interviewees commented on the simplicity of the rating, so we modified Wireframe A, **condensing the rating and detailed rating experience onto one screen**. Even though this design contains the same information as Wireframes A and C, since it is condensed onto one screen, the user does not feel as if they are completing multiple steps. In this design, the user would give ratings of the detailed categories, and then summarize their feelings with an overall rating at the bottom. The intention behind this layout is for the user to physically **see their ratings of each category and factor in the positive and negative aspects before leaving a final overall review**. With these four categories, the ambiguity problems from the five star rating system is addressed. Since the user sees a breakdown of their rating, they better understand their own overall rating, and the driver can see that as well.

Wireframe C

Our last design used the example of leaving a negative rating. **The user could not opt out of this one**, and was shown the overall rating first, then continued to give ratings for detailed categories.

This particular example only shows a negative experience, but the flow would be the same for a positive review.

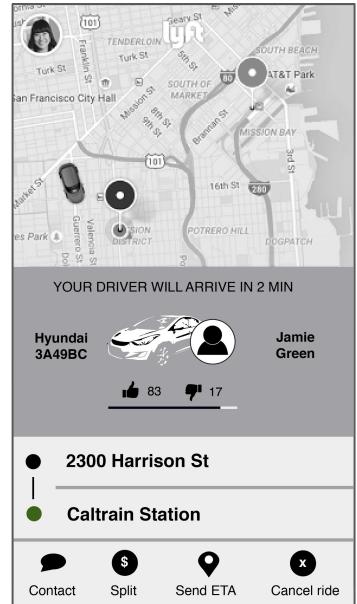
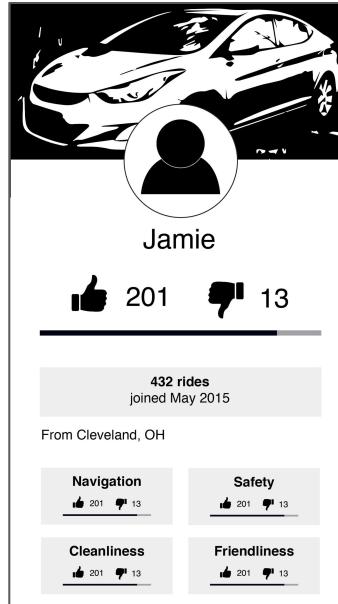
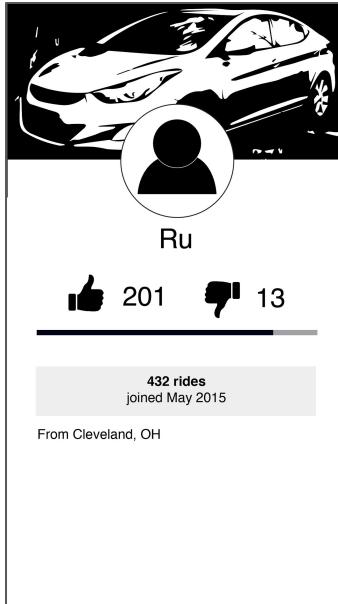
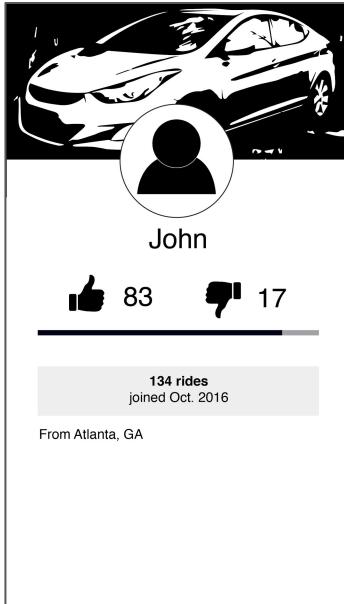
The image displays three wireframes illustrating a feedback process:

- Wireframe 1:** A driver profile for "Tom Neilson" is shown. Below the profile picture are five buttons for tipping: "No Tip", "\$1", "\$2", "\$5", and "Other". A "Payment" button is also present. At the bottom, it shows "\$11.90 credit applied" and the total amount "\$12.90".
- Wireframe 2:** A "Sorry!" message is displayed with a back arrow. It asks for feedback to provide a better experience next time. Below this are four rating categories with thumbs up and thumbs down icons:
 - Safety
 - Navigation
 - Friendliness
 - CleanlinessA text input field for comments and a "Submit" button are at the bottom.
- Wireframe 3:** A "Thank you" message is shown, featuring a decorative ribbon icon.

Wireframe C

This design is similar to Wireframe A, in which the user gives an overall rating before moving on to detailed ratings. However, **this design does not give the user an option to opt out of giving a positive or negative rating on the main screen. An overall review must be given to proceed.** The rationale behind this design stemmed from a current Lyft rating model in which users can “flag” a driver for one of the four categories, Safety, Navigation, Friendliness, and Cleanliness. This option is only shown during a negative review. This prevents users from leaving a negative review without giving a reason. We wanted to expand upon this idea by enabling the same system, but instead of just marking the negative aspects, giving feedback on what went well, despite the overall negative rating. This design decision is also informed in response to subjectivity of a “bad” rating discussed in interviews. By giving the user four specific options to declare **why something was bad in addition to just a “bad” rating** (in contrast to a 1, 2, or 3 star rating), there is more clarity.

Wireframe - driver profile



Map picture retrieved from: <https://help.lyft.com/hc/en-us/articles/213584088-How-to-get-picked-up-as-a-passenger>

Wireframe - driver profile

Primary interviews informed us that 3/4 of our interviewees **did not care much about the driver profile**, and had a preconception that **Lyft would be screening the drivers**. This conversation showed that Lyft passengers trust the company to have good drivers, and are not necessarily influenced by the average star rating of the drivers. Since we changed the five star system on the ratings interface, we also needed to update the five star system on the driver profile. We decided to show the exact number of overall thumbs up or thumbs down ratings (positive or negative) to the user, to eliminate the passenger's effort of interpreting an average rating. The rationale behind showing the user three different, but similar, driver profiles was to see if the number of overall ratings (the total of thumbs up and down), as well as the number of negative ratings, prompted the user to want to know more about the breakdown of why the driver was receiving negative reviews. This is why we included a driver profile with the individual tag ratings.

Final Design

Comparison

The current Lyft rating system vs. new design

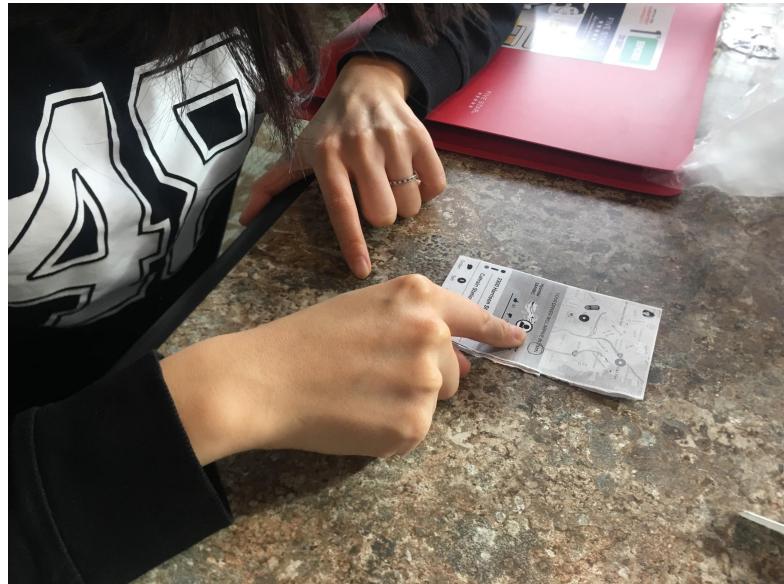
	The current Lyft rating system	New design
mechanism	Five-star rating	Thumbs up / Thumbs down / skip
Driver's perspective	Struggling in not achieving 4.6	Not overwhelming by strict deactivation standard
Rider's perspective	Confused with the meaning of 1-5 stars	Clear and effective

Usability testing

Method

We recruited 4 participants, and each of them participated in a usability testing session for around 20 minutes.

Participants were asked to **think aloud** while interacting with the system. The purpose of doing so is to observe if users are able to complete a pre-designed tasks with minimum frustrations. We paid attention to both participants articulations and facial emotions, and tried to identify what caused them to have frustrated moments.

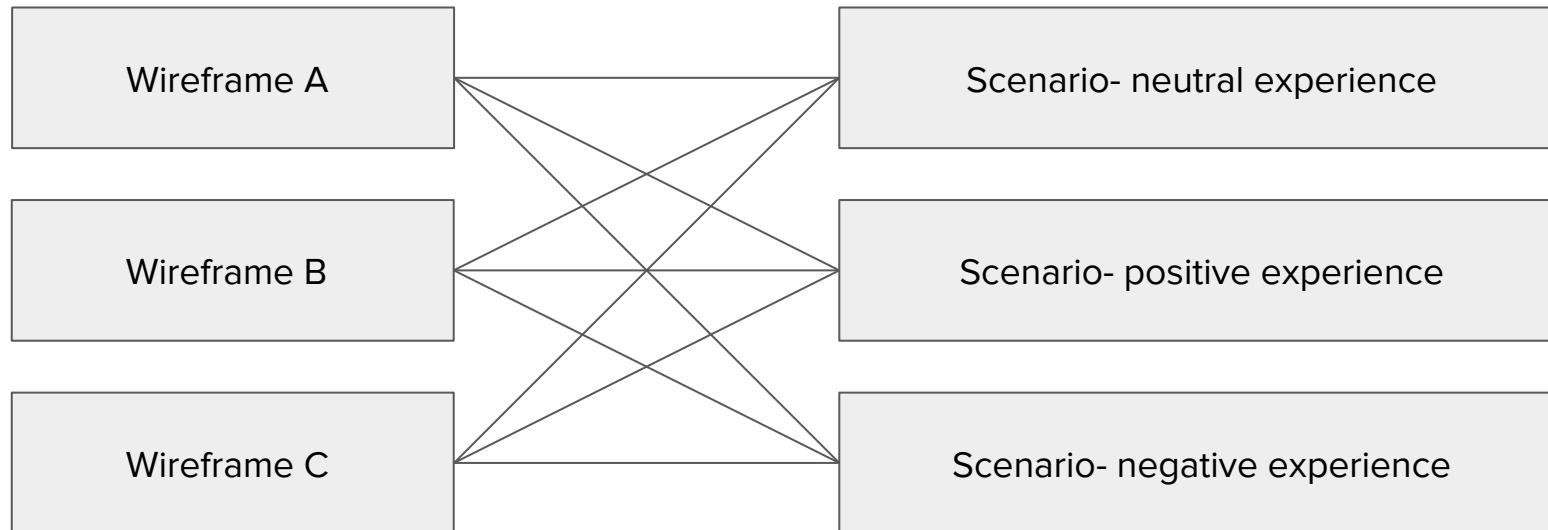


* Please see Appendix C for our usability testing protocol

Usability testing

Settings

We tested three scenarios for three wireframes. We tested three wireframes to analyze which task flow is the most easy to follow for the users. We covered three scenarios to test if our design can fulfill users' needs when they face different riding experiences.



Usability testing

Scenarios

We prompted users to behave normally as they usually interact with the system and imagine the situation they usually face when taking a Lyft. For example, rushing into an event or walking with lots of stuff in their hands.

Scenario 1: neutral experience

You just finished your trip with a lyft driver. You safely arrived at your destination, but you were not completely satisfied with your trip. You observed that the seats were full of stains and there was an unpleasant smell in the car. Even though the driver was very warm-hearted, he was so talkative that you felt a bit annoyed. The overall experience was OK, since the car arrived at the destination on time.

Scenario 2: positive experience

You had a wonderful lyft experience today. The driver was very nice and the car was very clean. The driver had free bottled water in the car for you to drink, and the music that he played was just about right. You was completely satisfied with your experience, and would like to leave a \$5 tip.

Scenario 3: negative experience

It was the worse ever lyft experience you had. The driver seemed to be under some type of influence and kept talking about things that were irrelevant to your trip. He deviated from the GPS's suggested route and it took forever to get to the destination. The actual cost was way over the estimated cost, and the driver did not help with your luggage at all.

Usability testing

Results & Insights

- 3 of 4 participants did not understand and did not prefer Wireframe A
- 2 of 4 participants were confused by the thumbs down icon being first, then thumbs up second, when reading left to right
- 4 of 4 participants were confused by the thumbs down and thumbs up icons on the bottom of the screen in Wireframes A and B, they were unaware that the icons were buttons, and pressed submit without touching the icons
- In general, there was resistance from participants when presented with multiple consecutive screens to give ratings
- In general, there was consensus among participants that Wireframe B was their preferred option

Iteration

In order to address the problems revealed through user testing, we continued to iterate upon our designs to create a final design. This design combined some elements from our initial wireframes, modified some, and discarded others.

- Loosely combine concepts of Wireframes A and B to make a one screen interaction, revealing detail without sacrificing detail
- Allowing the user to complete the rating in one step (two clicks minimum to exit rating and payment screen)
- Addressed a hierarchy issue in which users were confused about the relevance of an overall rating
- More clearly display thumbs up and thumbs down buttons
- Change the order of thumbs up and thumbs down buttons so they read positive to negative, left to right
- Provide a short text indication to the user that they should tap the thumbs up or thumbs down button to leave a rating

Final Design

Rating

This is the experience for leaving a **positive** rating.



Tom Neilson

No Tip \$1 \$2 \$5 Other

Payment VISA **3655

\$11.90 credit applied
\$12.90

Rate your experience.

Submit



Tom Neilson

No Tip \$1 \$2 \$5 Other

Payment

\$11.90 credit applied
\$12.90

Rate your experience.

Give us some detailed feedback. (Optional)

Safety Navigation

Friendliness Cleanliness

leave comments here ...

Submit



Tom Neilson

No Tip \$1 \$2 \$5 Other

Payment

\$11.90 credit applied
\$12.90

Rate your experience.

Give us some detailed feedback. (Optional)

Safety Navigation

Friendliness Cleanliness

leave comments here ...

Submit

Final Design

Rating

This is the experience for leaving a **negative** rating.



Tom Neilson

No Tip \$1 \$2 \$5 Other

Payment VISA **3655

\$11.90 credit applied
\$12.90

Rate your experience.

Submit



Tom Neilson

No Tip \$1 \$2 \$5 Other

Payment

\$11.90 credit applied
\$12.90

Rate your experience.

Give us some detailed feedback. (Optional)

Safety Navigation

Friendliness Cleanliness

leave comments here ...

Submit



Tom Neilson

No Tip \$1 \$2 \$5 Other

Payment

\$11.90 credit applied
\$12.90

Rate your experience.

Give us some detailed feedback. (Optional)

Safety Navigation

Friendliness Cleanliness

leave comments here ...

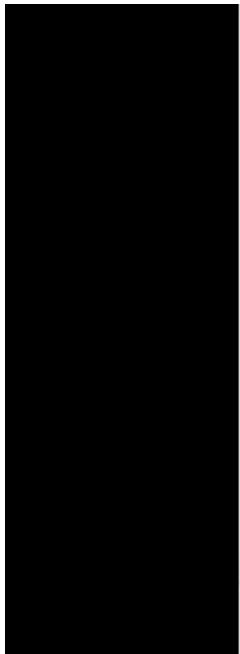
Submit

Final Design

Rating

The rationale behind the final ratings design is to clear up confusion discovered in usability testing. Each issue with the wireframes used in usability testing was revisited in order to improve the experience. Changes ranged from button placement and hierarchy, to simplifying interactions and screens. Overall, users wanted a simplified experience, in which they could give an overall rating and feedback on one screen. This final ratings design addresses our problem statements of an inefficient and ambiguous rating system, as well as a limited amount of feedback to the driver.

Final Design - Animation



Tom Neilson

No Tip \$1 \$2 \$5 Other

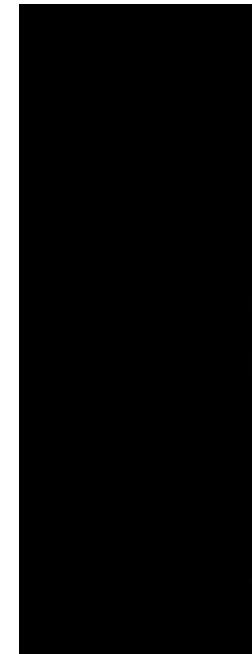
Payment

\$11.90 credit applied
\$12.90

Rate your experience.

Give us some detailed feedback. (Optional)

Submit

The user interface for a positive rating. It shows a profile picture of 'Tom Neilson', a tip amount selection bar, a payment input field, a total amount of '\$12.90' with a note about credit applied, and a rating section with a green thumbs-up icon and a red thumbs-down icon. There is also an optional feedback section and a 'Submit' button at the bottom.

\$12.90

Rate your experience.

Give us some detailed feedback. (Optional)

Safety

Friendliness

Navigation

Cleanliness

leave comments here ...

Submit

The user interface for a positive rating, identical to the first frame but with a different background color. It includes a total amount of '\$12.90', a rating section with icons, an optional feedback section, and a 'Submit' button.

Positive Rating Flow

Link: <https://drive.google.com/open?id=0Bx130n0eG-rhS0pYQkk1ZmhIS2M>

Negative Rating Flow

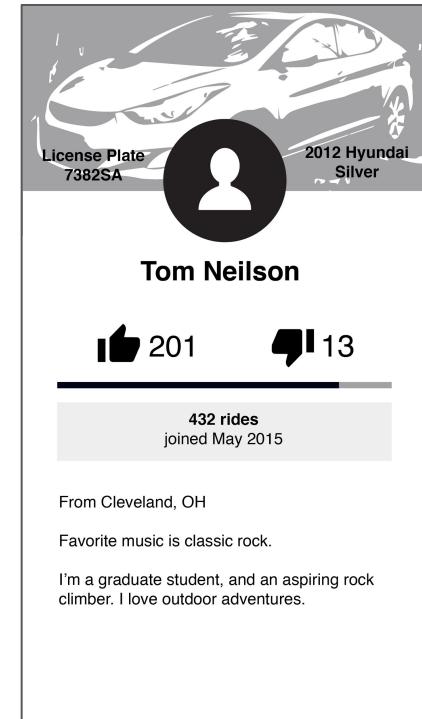
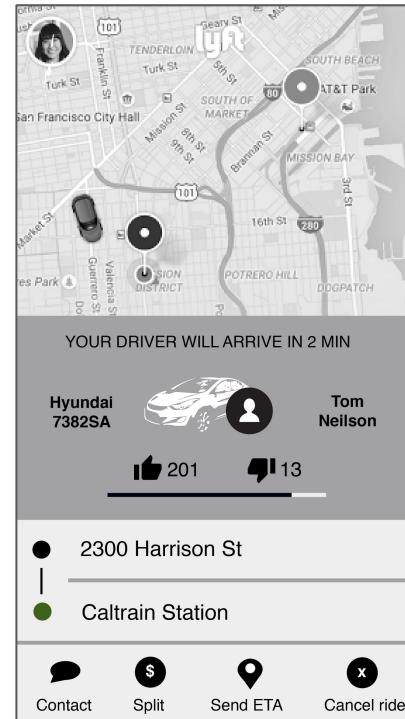
Link: <https://drive.google.com/open?id=0Bx130n0eG-rheIBVzFuSzdBUVk>

Final Design

Driver profile

This is the experience shown after requesting a Lyft. The first screen illustrates quick, critical information about the driver. The second screen would be shown if the passenger would like to see a more detailed driver profile.

These designs improve upon the existing Lyft driver profiles.



Final Design

Driver profile

Not much changed in the final design of the driver profile and expanded driver profile after usability testing. This was in part due to this aspect of the overall design not being emphasized as much in testing, and also because it's most relevant feature to this project, the rating itself, was received positively in usability testing. A design decision was made to not show a breakdown of tags in the expanded driver profile because there was no strong evidence from usability testing to include it in the final design. In general, participants liked the thumbs up and thumbs down bar, as well as the number of positive and negative ratings. The final design for both screens shown is modeled after the current Lyft driver profile, with the exception of the rating system replacement.

Lyft Rating Video

This video shows a simplified explanation of our final design prototype. It outlines why the thumbs up, thumbs down system is more helpful to the users, and how the tagging system addresses passengers' changing needs.

Video link: <https://youtu.be/3id3dje9l1q>



Reflection

Although our final designs took the feedback from users during usability testing and address our problem statements of an inefficient and ambiguous rating system, as well as a limited amount of feedback to the driver, there is room to improve upon the design. Our designs could be further tested to show confusing points on the interface, as well as for interaction clarity. Reflecting on the designs as a group, we discussed how perhaps the tags themselves could be changed for clarity; maybe four categories is too few, or maybe they could be different categories. We also discussed the limitations of a thumbs up and thumbs down system; although two options allow users to give a concise reaction, perhaps there is not enough gray area for users who have a neutral experience. We hope gain to gain feedback to improve upon this problem in the future.

Conclusion

We discovered strengths and weaknesses of different ratings systems through secondary research, which informed our interview questions and enabled us to explore Lyft's rating system further. Our primary research began by interviewing users about their past experiences with Lyft, as well as their preferences and frustrations about the current interface. After organizing interview trends into an affinity map, we analyzed trends in frustrations and preferences in the context of the whole ride experience; before the ride, during the ride, and after the ride. Our initial sketches and wireframes attempted to solve these problems encountered, but usability testing made the weaknesses in our designs apparent. Users were confused about placement of elements and the progression of screens. Users showed little interest in learning more about the driver profile, and were satisfied with a simple interface. Our final designs took the feedback from users during usability testing. The final design addresses our problem statements of an inefficient and ambiguous rating system, as well as a limited amount of feedback to the driver. This design still has room for improvement, and could be subject to future usability testing and analysis.

References

- Allen, C. (2006). Using 5-Star Rating Systems. Retrieved from http://www.lifewithalacrity.com/2006/08/using_5star_rat.html
- Anderson, M. (2014). 88% Of Consumers Trust Online Reviews As Much As Personal Recommendations. Retrieved from <http://searchengineland.com/88-consumers-trust-online-reviews-much-personal-recommendations-195803>.
- Cook, J. (2015). Uber's internal charts show how its driver-rating system actually works. Business Insider. Retrieved from <http://www.businessinsider.com/leaked-charts-show-how-ubers-driver-rating-systemworks-2015-2>.
- Etzioni, A. (2017). Cyber Trust.
- Hanington, B., & Martin, B. (2012). Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Rockport Publishers.
- Hu, N., Pavlou, P. A., & Zhang, J. (2006). Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication. In Proceedings of the 7th ACM conference on Electronic commerce(pp. 324-330). ACM.
- Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. Communications of the ACM, 52(10), 144-147.
- Lyft Inc. (2017). Driver and passenger ratings. Retrieved from <https://help.lyft.com/hc/en-us/articles/213586008-Driver-and-passenger-ratings#define>.
- Rainie, L. & Hitlin, P. (2004). Use of Online Rating Systems. Retrieved from <http://www.pewinternet.org/2004/10/20/use-of-online-rating-systems/>.
- Smith, A. (2016). Shared, Collaborative and On Demand: The New Digital Economy. Retrieved from <http://www.pewinternet.org/2016/05/19/on-demand-ride-hailing-apps/>.
- Sparling, E. Isaac, and Shilad Sen. "Rating." Proceedings of the fifth ACM conference on Recommender systems - RecSys 11, 2011, doi:10.1145/2043932.2043961.

Appendix A - interview protocol

Preamble

Thank you for participating in our project. This project is a partial fulfillment of the requirements for the completion of the CGT572 (*Special Topics in Human-Centered Design & Development*) course offered by Purdue Polytechnic Institute.

We're not evaluating you in any way, but we're learning from you about how you feel about Lyft's five-star rating system. With this, we hope we can make our product better for people like you. There are no right or wrong answers to any of the questions I'm asking in this study - We are simply interested in understanding your thoughts and feelings about the system. We will conduct this study mostly as a discussion - I will ask you questions to better understand what you do.

Here's how the session's going to be scheduled. We'll ask you a few general questions about how you exercise. Then we'll go through a few questions regarding your opinions towards fitness applications. Feel free to say anything that comes to your mind; don't worry about offending us since we didn't design this product.

We will record a little video of you so that I can go back and review things later and make sure we get everything right. We won't use your name in connection with the recordings or the results. The videotapes will only be used internally and never shared anywhere with anyone.

Should you have any questions, please feel free to ask me at this point.

Appendix B - interview questions

Personal Info

- Gender
- Age
- Occupation

Riding experience

- What's the most frequent type of transportation do you use? How often?
- Under what circumstances do you usually use Lyft (or other ride-sharing)? How often?
- What's your impression on Lyft?
- Do you check the driver's profile before taking Lyft? What information you look for?
- What do you usually do after you get in the car? Describe a typical journey you had.
- Describe a good lyft experience you have had. Why was it good?
- Describe a bad lyft experience. Why was it bad? What rating did you give to that driver?
- What do you think of a driver with 4.6, 4.8, 5 stars? Would you feel differently or act differently when you taking the ride?
- What do you care the most when taking Lyft?
- Have you ever been asked to rate 5 stars by the driver?

Current Interface

- What do you think of the rating system? (show them the interface of a current interface)
- What features (designs) do you like best? Why?
- What aspects do you want to improve? You can make modifications on the paper.
- Are there anything you wish to have in the system, but not currently available?
- Do you ever rate other online services? (Amazon purchases, apps, etc.)
- What influences your driver rating? (e.g. arrival time, professionalism, car quality, cleanliness, friendliness etc.)
- Do you rate your drivers all the time? If not, why?
- Can you explain your understanding of 1-5 star?
- Do you know how your rate affects the drivers? / Do you know the purpose of rating?

Appendix C - usability testing protocol

Usability testing: participants' understanding of different rating UI ideas

Wireframe A. Thumbs up, neutral, Thumbs down (Overall rating → detailed rating)

Wireframe B. Thumbs up, neutral, Thumbs down (Detailed rating → overall rating)

Wireframe C. Thumbs up, Thumbs down

Task: Rate your Riding Experience

Subtask 1: Give tips

Subtask 2: Overall rating (Thumbs up / Thumbs down)

Subtask 3: Detailed Rating

Prompted users to behave normally as they usually interact with the system and imagine the situation they usually face when taking a Lyft. For example, rushing into an event or walking with lots of stuff in their hands.

Task 1 (Neutral)

You just finished your trip with a lyft driver. You safely arrived at your destination, but you were not completely satisfied with your trip. You observed that the seats were full of stains and there was an unpleasant smell in the car. Even though the driver was very warm-hearted, but you felt he was so talkative that you felt a bit annoyed. The overall experience was OK, since the car arrived at the destination on time.

Task 2 (Positive)

You had a wonderful lyft experience today. The driver was very nice and the car was very clean. The driver had free bottled water in the car for you to drink, and the music that he played was just about right. You was completely satisfied with your experience, and would like to leave a \$5 tip.

Task 3 (Negative)

It was the worse ever lyft experience you had. The driver seemed to be under some type of influence and kept talking about things that were irrelevant to your trip. He deviated from the GPS's suggested route and it took forever to get to the destination. The actual cost was way over the estimated cost, and the driver did not help with your luggage at all.

A/B testing: participants' preferences about driver's profile

We test

1. If the preview profile screen is clear enough
2. if different types of data about driver's profile affect participants' willingness to look for more detailed information.