
Mixed methods for evaluating user satisfaction

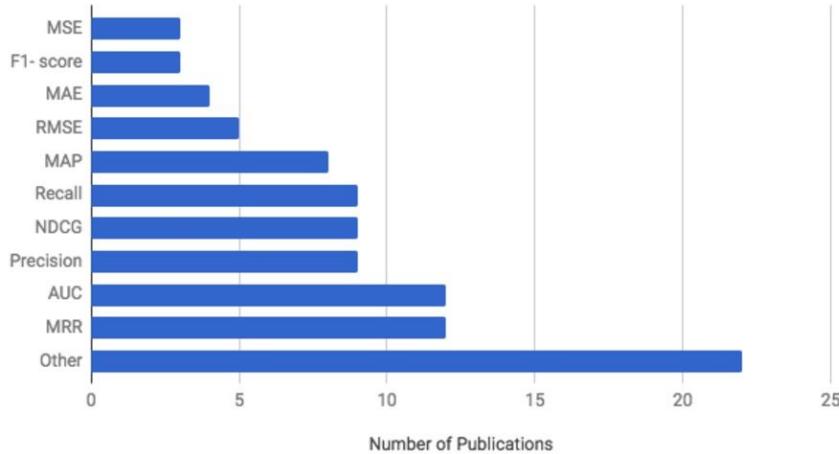
Jean Garcia-Gathright 
Christine Hosey 
Brian St. Thomas 
Ben Carterette 
Fernando Diaz 

RecSys '18
October 2, 2018
Vancouver, BC



How do we evaluate recommender systems today?

Evaluation metrics in use



- System-oriented metrics
- Easy to compute if there is abundant behavioral data
- Missing the user perspective

How do we evaluate recommender systems today?

- Less common: user studies
- Includes the users' perspective, but at a smaller scale

movielens

List A (10 movies) List B (10 movies)

| Movie A | Movie B |
|--|--|
| Pépé le Moko 1937 94 min Action, Crime | Fear City: A Family-Sty 1994 93 min Comedy |
| The Mummy's Curse 1944 52 min Horror | Connections (1978) 1977 |
| Terra Libertad 1994 109 min Drama, History | Ween: Live in Chicago 2004 120 min |
| Children of Paradise 1945 190 min Drama, Romance | Hellhounds on My Trail |
| What Time Is It There? 2000 116 min Drama, Romance | Heimat: A Chronicle of 1984 925 min |

Survey (25 questions)

Lists A and B contain the top movie recommendations for you from different "recommenders". Please answer the following questions to help us understand your preferences about these recommenders.

1. Based on your first impression, which list do you prefer?

| Much more A than B | About the same | Much more B than A |
|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

2. Which list has more movies that you find appealing?

| Much more A than B | About the same | Much more B than A |
|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

3. Which list has more movies that might be among the best movies you see in the next year?

| Much more A than B | About the same | Much more B than A |
|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

4. Which list has more obviously bad movie recommendations for you?

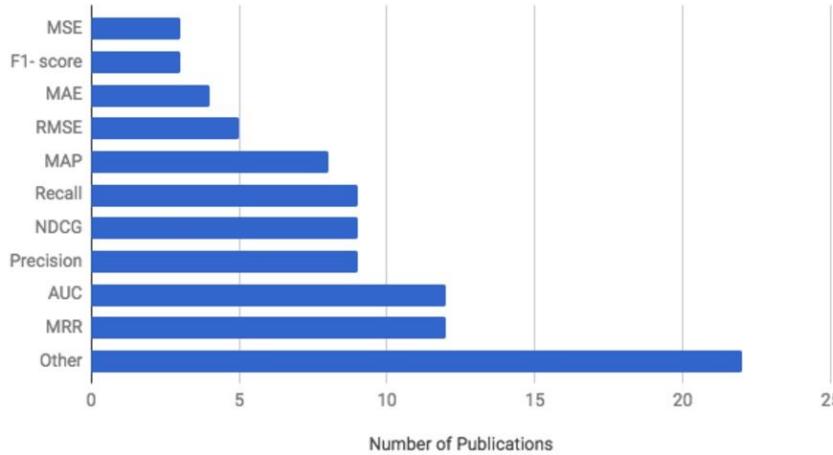
| Much more A than B | About the same | Much more B than A |
|-----------------------|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

scroll down for more scroll down for more (why so many questions?)

Ekstrand, M.D., Harper, F.M., Willemsen, M.C. and Konstan, J.A. User perception of differences in recommender algorithms. RecSys '14.

These approaches tend to be applied independently

Evaluation metrics in use



movielens

List A (10 movies)



Pépé le Moko
1937 94 min
Action, Crime

List B (10 movies)



Fear City: A Family-Style
1994 93 min
Comedy



The Mummy's Curse
1944 62 min
Horror



Connections (1978)
1977



Terra Libertad
1994 109 min
Drama, History



Ween: Live in Chicago
2004 120 min



Children of Paradise
1945 190 min
Drama, Romance



Hellhounds on My Trail



What Time Is It There?
2000 116 min
Drama, Romance



Heimat: A Chronicle of
Germany
1984 925 min

Survey (25 questions)

Lists A and B contain the top movie recommendations for you from different "recommenders". Please answer the following questions to help us understand your preferences about these recommenders.

1. Based on your first impression, which list do you prefer?

Much more A than B About the same Much more B than A

2. Which list has more movies that you find appealing?

Much more A than B About the same Much more B than A

3. Which list has more movies that might be among the best movies you see in the next year?

Much more A than B About the same Much more B than A

4. Which list has more obviously bad movie recommendations for you?

Much more A than B About the same Much more B than A

scroll down for more

scroll down for more (why so many questions?)

Mixed methods is a way to robustly evaluate user satisfaction

- Qualitative research can help us understand the user perspective on satisfaction
- Combine with quantitative behavioral data at scale to get a more complete understanding



Outline of the tutorial

| | | | |
|------------|---|---|--|
| 9:00-10:30 | Overview of mixed methods | | |
| | | Qualitative | Quantitative |
| | Data Collection | <ul style="list-style-type: none">• Interviews• Surveys | <ul style="list-style-type: none">• Implicit feedback from log data |
| | Data Analysis | <ul style="list-style-type: none">• Inductive analysis• Deductive analysis | <ul style="list-style-type: none">• Exploratory analysis• Satisfaction modeling |
| | Best practices for mixed methods teams | | |
| | Hypothesis Testing | | <ul style="list-style-type: none">• Experiment design• Statistical tests |

Overview of mixed methods research

- What is mixed methods and why is it useful?
 - Mixed methods study designs
-

What is mixed methods and why is it useful?

“[In] mixed methods research:

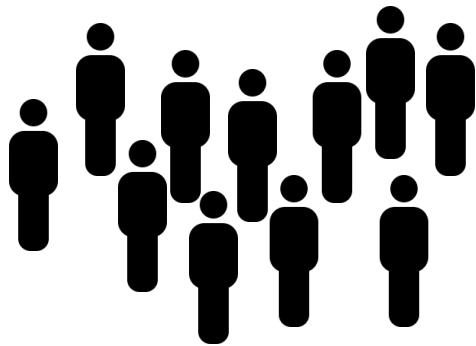
a researcher or team of researchers

combines elements of qualitative and quantitative
approaches

for breadth and depth of understanding and
corroboration.”

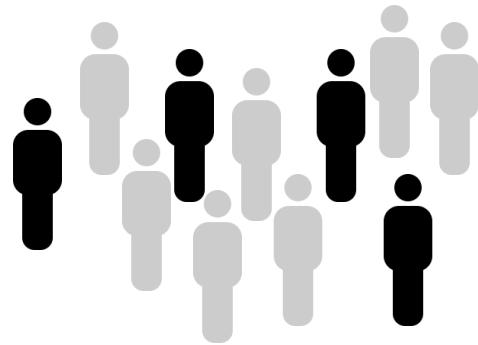
Johnson, R.B., Onwuegbuzie, A.J. and Turner, L.A., 2007. Toward a definition of mixed methods research. Journal of mixed methods research, 1(2), pp.112-133.

users





qualitative
research



Qualitative methods

UNDERSTAND
NEEDS

REFINE
HYPOTHESIS

EVALUATE
CONCEPTS

EVALUATE
INTERFACES

EVALUATE
EXPERIENCE

Interviews

Co-creation

Intercepts

Ethnography

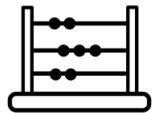
Focus groups

Usability testing

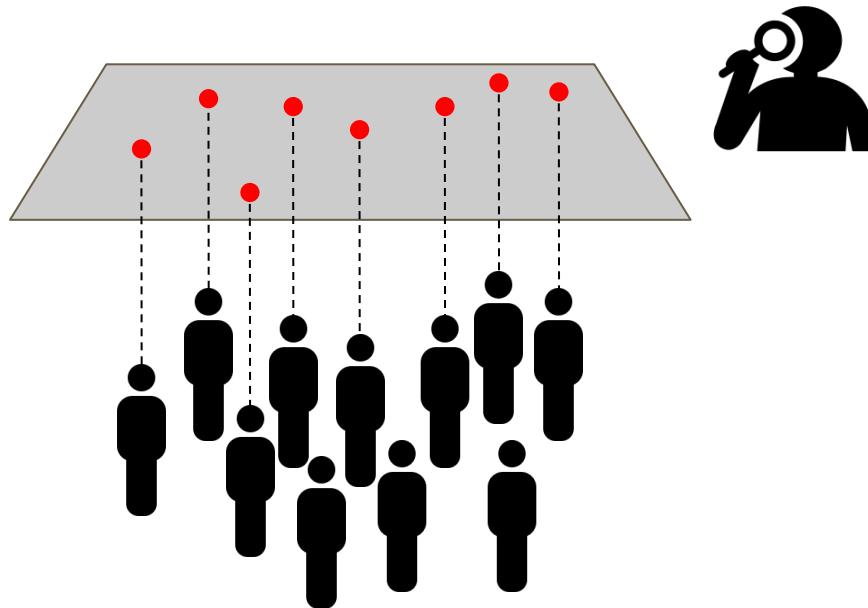
Surveys

Card sorting

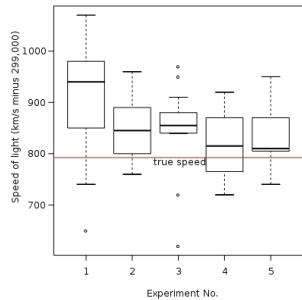
Diary research



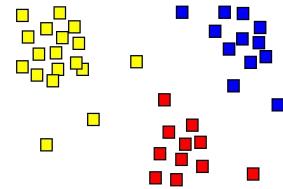
quantitative
research



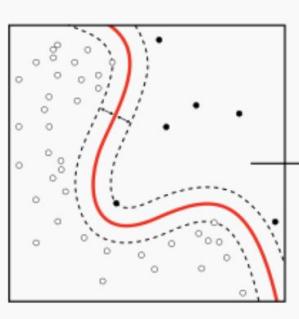
Quantitative methods



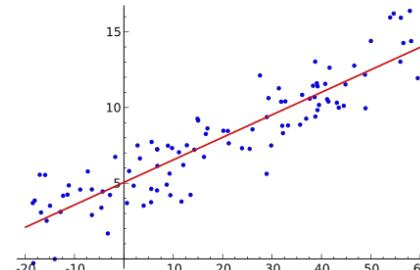
Descriptive statistics - summarize the data



Clustering - group together similar data points

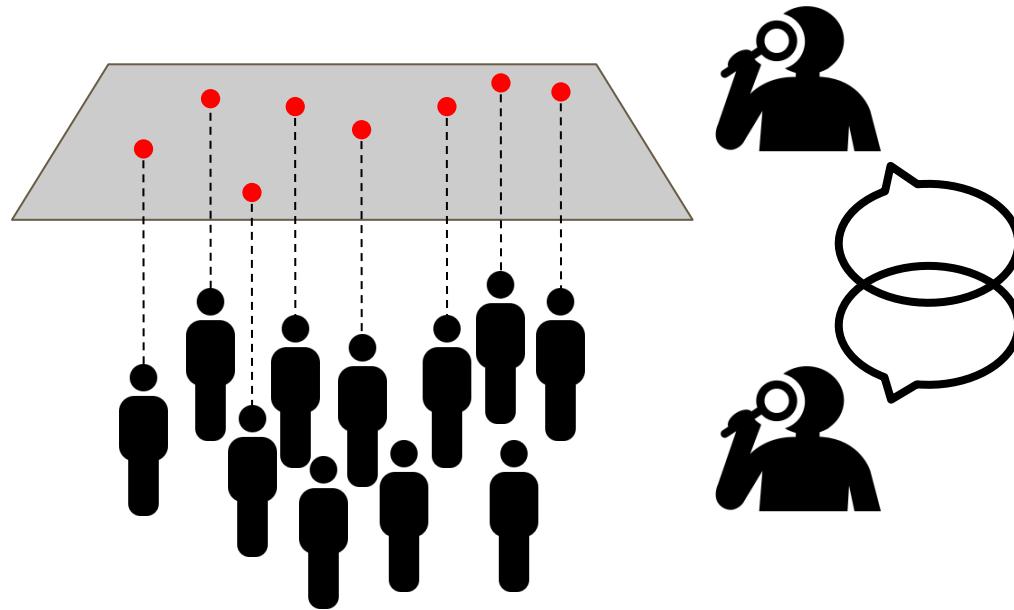


Classification - place data points into categories

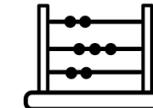


Regression - find relationships between variables

mixed methods
research



Qualitative and quantitative approaches are complementary

- Qualitative 
 - Small sample size
 - Emphasize the voices of the participants
 - Explore meaning and understanding of constructs
- Quantitative 
 - Large sample size
 - Generalize to a population
 - Assess magnitude and frequency of constructs

When should we use mixed methods?

- To augment experiments by incorporating the **perspectives of individuals**
- To **explain** initial quantitative results
- To qualitatively **explore** questions, variables, and theories prior to a quantitative study
- To obtain more **complete and corroborated** results

Creswell, J.W. and Clark, V.L.P. (2017). 'The Nature of Mixed Methods Research.'

In *Designing and conducting mixed methods research*. Sage publications.

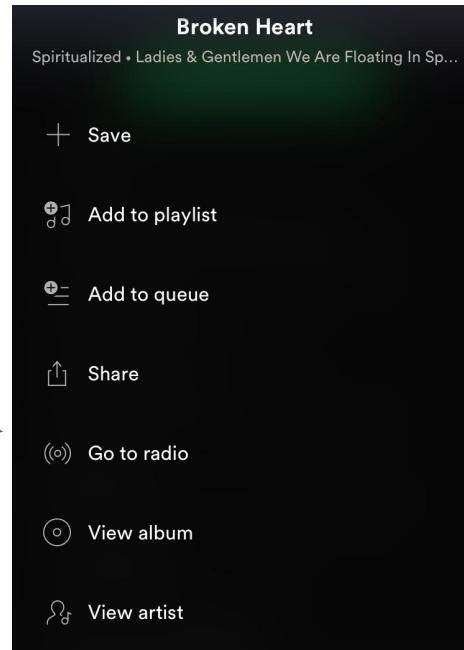
Mixed methods study designs

Mixed methods is
more than the sum of its parts.

We achieve this by **integrating** qualitative
and quantitative approaches.

Fetters, M.D. and Freshwater, D., 2015. The $1 + 1 = 3$ Integration Challenge.
Journal of Mixed Methods Research, 9(2), pp. 115 - 117.

Running example: Discover Weekly

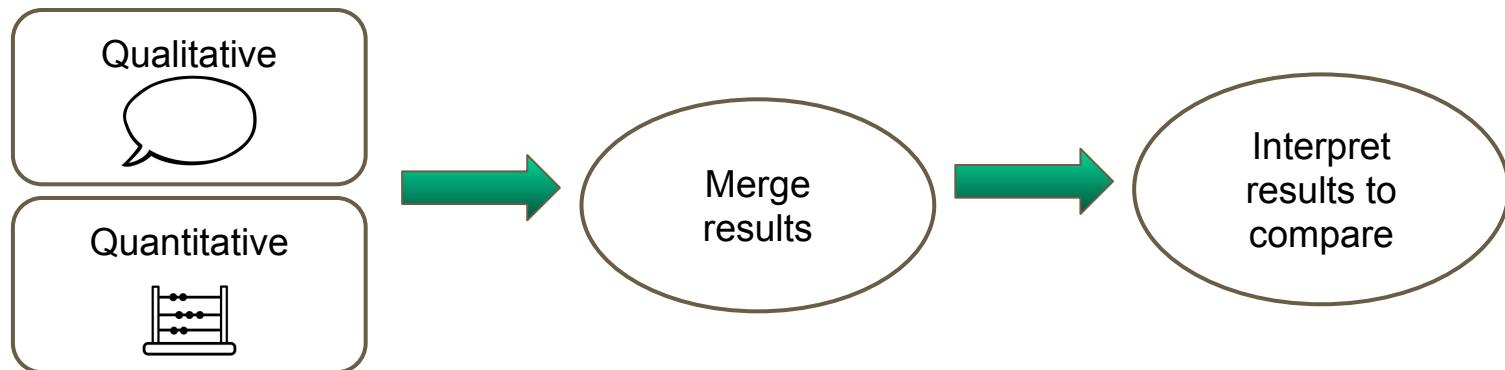


- 30 unfamiliar tracks
- Personalized to your taste
- Updated every Monday
- Tracks can be interacted with in several ways

Running example: Discover Weekly

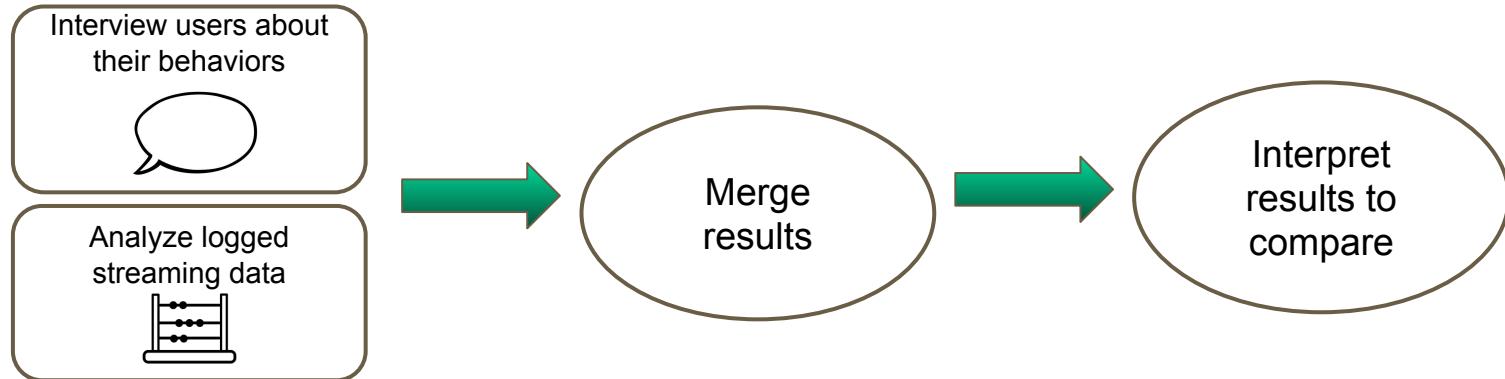
- Understand users' expectations
- Determine which interactions signal satisfaction
- Evaluate performance of the recommender system

Convergent design



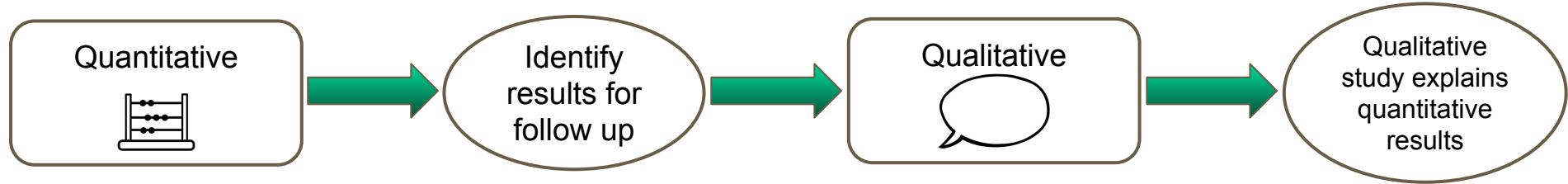
Creswell, J.W. and Creswell, J.D. (2017). 'Mixed Methods Procedures.' In Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Discover Weekly: is listening time a good proxy for satisfaction?



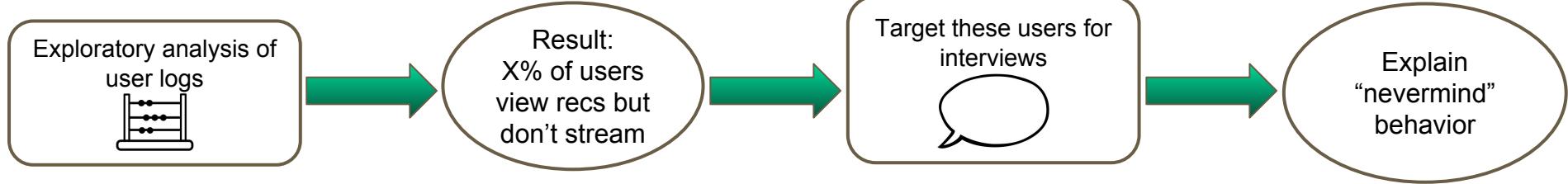
Creswell, J.W. and Creswell, J.D. (2017). 'Mixed Methods Procedures.' In Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Explanatory sequential



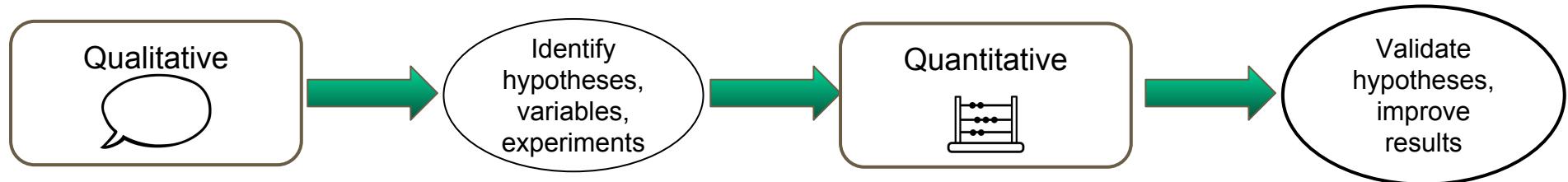
Creswell, J.W. and Creswell, J.D. (2017). 'Mixed Methods Procedures.' In Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Discover Weekly: why do some users say “nevermind”?



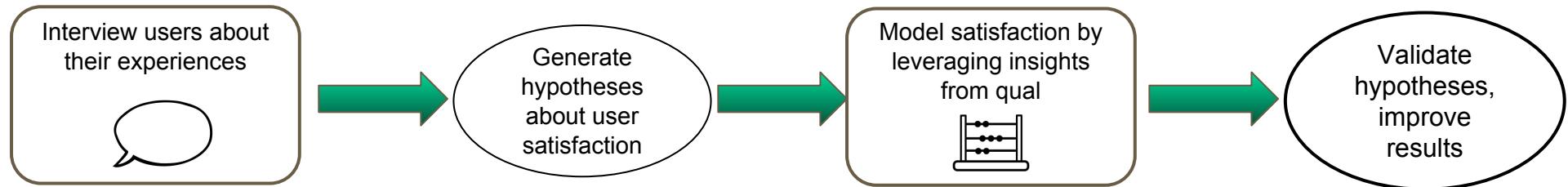
Creswell, J.W. and Creswell, J.D. (2017). 'Mixed Methods Procedures.' In Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Exploratory sequential



Creswell, J.W. and Creswell, J.D. (2017). 'Mixed Methods Procedures.' In Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Discover Weekly: what is user satisfaction, and how do we measure it?



Creswell, J.W. and Creswell, J.D. (2017). 'Mixed Methods Procedures.' In Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Data collection

- Qualitative
 - Interviews
 - Surveys
 - Quantitative
 - Attention
 - Interaction
 - Satisfaction
 - Retention
-

Qualitative Data Collection

Qualitative methods

UNDERSTAND
NEEDS

REFINE
HYPOTHESIS

EVALUATE
CONCEPTS

EVALUATE
INTERFACES

EVALUATE
EXPERIENCE

Interviews

Co-creation

Intercepts

Ethnography

Focus groups

Usability testing

Surveys

Card sorting

Diary research

Qualitative methods

UNDERSTAND
NEEDS

REFINE
HYPOTHESIS

EVALUATE
CONCEPTS

EVALUATE
INTERFACES

EVALUATE
EXPERIENCE

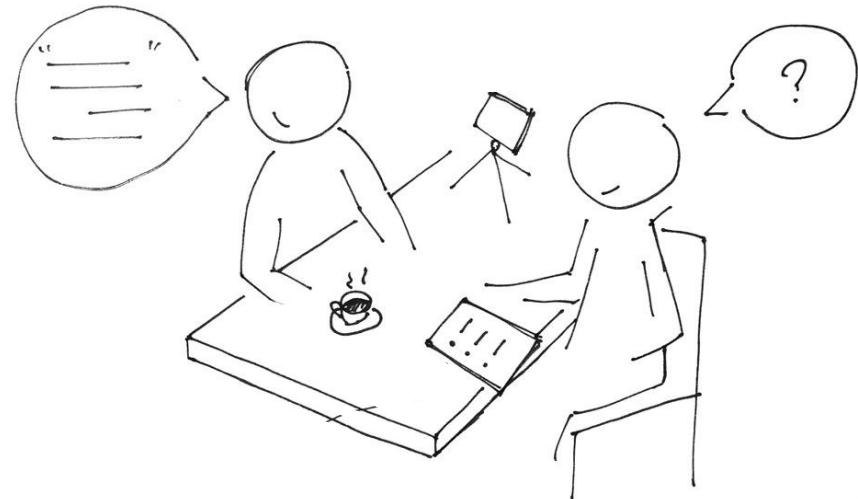
Interviews

Surveys

Interviews

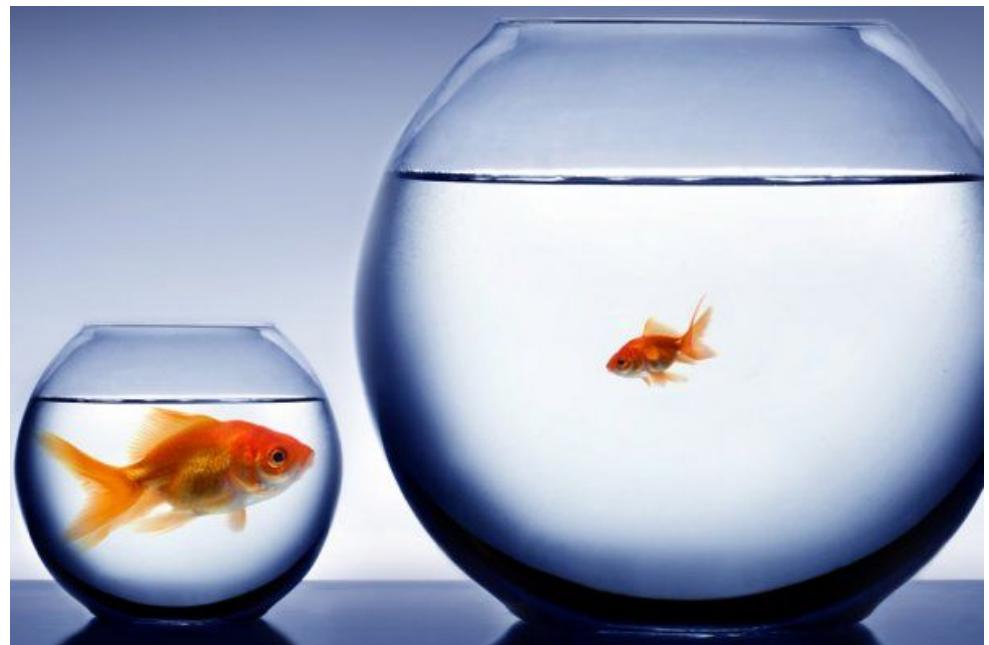
What is a semi-structured interview?

An interview in which the researcher uses a **guide** with questions and topics that must be covered, but has some **discretion** about the order in which questions are asked and has **flexibility** to cover unscripted material.



Why use an interview?

- Builds a **holistic** snapshot
- Allows people to express their **own thoughts and feelings**
- Uncovers what is **not directly observable**



Selecting participants

Goal: Represent wide range of users

(Not the goal: To be able to directly compare between groups)

Participant selection

- **Define cohorts:** levels of representation that you want to built into your recruiting strategy
- **Include a mix:** other dimensions you care about representing (e.g., gender, age, behaviors) across cohorts
- **Recruit:** 3-5 participants per cohort (expect cancellations and no-shows)



Basic interview structure

Introduction

Provide context:

- + Introduce yourself
- + Purpose and structure
- + Disclosures (recording, confidentiality)
- + Offer to answer questions

Opening

Ease participants into the process:

- + Start easy
- + Nothing too personal or difficult to answer

Content

Have a conversation:

- Funnel
- + Start broad
- + Get increasingly specific

Closing

Appreciate your participants:

- + Allow for last thoughts or questions
- + Thank them
- + Compensate (previously agreed upon)

Basic interview structure

Introduction

Provide context:

- + Introduce yourself
- + Purpose and structure
- + Disclosures (recording, confidentiality)
- + Offer to answer questions

Opening

Ease participants into the process:

- + Start easy
- + Nothing too personal or difficult to answer

Content

Have a conversation:

- Funnel
- + Start broad
 - + Get increasingly specific

Closing

Appreciate your participants:

- + Allow for last thoughts or questions
- + Thank them
- + Compensate (previously agreed upon)

Basic interview structure

Introduction

Provide context:

- + Introduce yourself
- + Purpose and structure
- + Disclosures (recording, confidentiality)
- + Offer to answer questions

Opening

Ease participants into the process:

- + Start easy
- + Nothing too personal or difficult to answer

Content

Have a conversation:

- Funnel
- + Start broad
 - + Get increasingly specific

Closing

Appreciate your participants:

- + Allow for last thoughts or questions
- + Thank them
- + Compensate (previously agreed upon)

Basic interview structure

Introduction

Provide context:

- + Introduce yourself
- + Purpose and structure
- + Disclosures (recording, confidentiality)
- + Offer to answer questions

Opening

Ease participants into the process:

- + Start easy
- + Nothing too personal or difficult to answer

Content

Have a conversation:

- Funnel
- + Start broad
- + Get increasingly specific

Closing

Appreciate your participants:

- + Allow for last thoughts or questions
- + Thank them
- + Compensate (previously agreed upon)

Interview best practices



Remain neutral

- + No leading questions
- + Focus on past experiences rather than future projections



Go deeper

- + Probes ("tell me more" "give an example" "why")
- + Get comfortable with silence
- + Allow for tangents



Don't be a jerk

- + No interrupting
- + No assuming - paraphrase to test understanding
- + Jog their memory (but only later)

Interview best practices



Remain neutral

- + No leading questions
- + Focus on past experiences rather than future projections



Go deeper

- + Probes (“tell me more” “give an example” “why”)
- + Get comfortable with silence
- + Allow for tangents



Don't be a jerk

- + No interrupting
- + No assuming - paraphrase to test understanding
- + Jog their memory (but only later)

Interview best practices



Remain neutral

- + No leading questions
- + Focus on past experiences rather than future projections



Go deeper

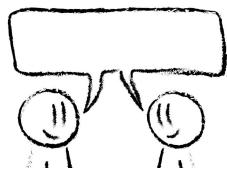
- + Probes (“tell me more” “give an example” “why”)
- + Get comfortable with silence
- + Allow for tangents



Don't be a jerk

- + No interrupting
- + No assuming - paraphrase to test understanding
- + Jog their memory (but only later)

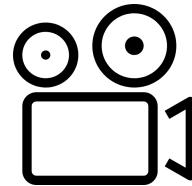
Preparing for and conducting the interview



Prepare & practice

It should feel like a conversation:

- + Don't simply read from a script
- + Practice on a real person



Record

Record participant and any activity (e.g., on phone) they will do:
+ Expressions, tone, actions are important data

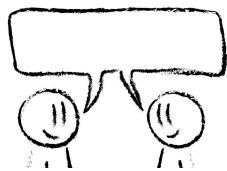


Transcribe

Send audio to a transcription service:

- + Aids in coding
- + Aids in capturing verbatim quotes

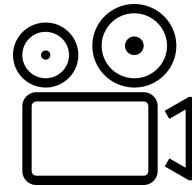
Preparing for and conducting the interview



Prepare & practice

It should feel like a conversation:

- + Don't simply read from a script
- + Practice on a real person



Record

Record participant and any activity (e.g., on phone) they will do:
+ Expressions, tone, actions are important data

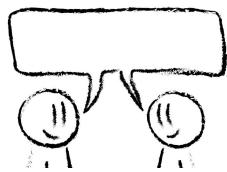


Transcribe

Send audio to a transcription service:

- + Aids in coding
- + Aids in capturing verbatim quotes

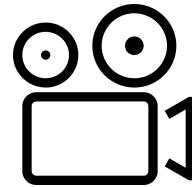
Preparing for and conducting the interview



Prepare & practice

It should feel like a conversation:

- + Don't simply read from a script
- + Practice on a real person



Record

Record participant and any activity (e.g., on phone) they will do:
+ Expressions, tone, actions are important data



Transcribe

Send audio to a transcription service:

- + Aids in coding
- + Aids in capturing verbatim quotes

Why use interviews to study user satisfaction?

Why use interviews to study user satisfaction?

Because if we don't ask people what makes them happy, we will make assumptions.

Example: Discover Weekly

What do we think might be proxies for user satisfaction?

Example: Discover Weekly

What do we think might be proxies for user satisfaction?

Listening
time?

Example: Discover Weekly

What do we think might be proxies for user satisfaction?

Listening
time?

Returning
next week?

Example: Discover Weekly

Listening time?

Example: Discover Weekly

**Listening
time?**



Listening time often has nothing to do with the quality of the experience.

Example: Discover Weekly

Listening time?



Listening time often has nothing to do with the quality of the experience.

Seek & save

"I'll hit the beginning of the song and I'll give it about 15 seconds to see if it catches me and if I think it's kind of interesting, then I'll tab in like maybe halfway into the song and listen to the middle of it and kind of just see what that sounds like, and if those two bits catch me, then I put it on the list."

— P10

Example: Discover Weekly

Listening time?



Listening time often has nothing to do with the quality of the experience.

Seek & save

*"I'll hit the beginning of the song and I'll give it about 15 seconds to see if it catches me and if I think it's kind of interesting, then I'll tab in like maybe halfway into the song and listen to the middle of it and kind of just see what that sounds like, and **if those two bits catch me, then I put it on the list.**"*

— P10

Artist exploration

*"I take stock of the ones I liked, **go research that artist** a little bit more. Kind of get lost down that **rabbit hole.**"*

— P4

*"If it is something that I really like, I will immediately stop what I am doing and **go listen the album** that whatever song came from. Then, that will usually get me on a **tangent** where I will just listen to that artist for the **rest of the day or for the next couple of days.**"*

— P7

Example: Discover Weekly

**Returning
next week?**

Example: Discover Weekly

**Returning
next week?**



Returning next week often does not signal a good experience this week.

Example: Discover Weekly

Returning next week?



Returning next week often does not signal a good experience this week.

Habit

“Then I just see it there on Monday and I click play on the big album box.”

— P1

*“Usually **Monday** mornings I’ll open it.”*

— P3

Example: Discover Weekly

Returning next week?



Returning next week often does not signal a good experience this week.

Habit

“Then I just see it there on Monday and I click play on the big album box.”

— P1

*“Usually **Monday** mornings I’ll open it.”*

— P3

In (or out of) a “discovery” mood

*“When there’s **a stretch of time** when I’m really in the mood to **discover music** and make playlists, then I might be more invested and interested and I can’t wait for Sunday and get my Discover Weekly.”* — P5

Example: Discover Weekly

Returning next week?



Returning next week often does not signal a good experience this week.

Habit

“Then I just see it there on **Monday** and I click play on the big album box.”

— P1

“Usually **Monday** mornings I’ll open it.”

— P3

In (or out of) a “discovery” mood

“When there’s **a stretch of time when I’m really in the mood to discover music** and make playlists, then I might be more invested and interested and I can’t wait for Sunday and get my Discover Weekly.” — P5

Bad algorithm

“If the first two songs on Discover Weekly aren’t that great, I’ll be like, “**This is a tough week for the algorithm** or whatever.” ... I might stop and listen to normal stuff and then see next week.”

— P9

Interviewing to build satisfaction hypotheses

Interviewing to build satisfaction hypotheses

Introduction

- Provide context:
 - + Introduce yourself
 - + Purpose and structure
 - + Disclosures (recording, confidentiality)
 - + Offer to answer questions

Ease into t
+ Star
+ Not
pers
to a

Content

Have a conversation:

- + Habits within the domain
- + Preferences within the domain
- + Attitudes toward the system
- + Reasons for system use
- + Habits with the system
- + Good experiences with the system
- + Bad experiences with the system
- + Behavior deep dive
 - Have you done X? If yes, why? If no, why not?

sing

e your
ts:

last
or

\$
em

ate
y agreed

Example: Discover Weekly

Sampling

Cohorts defined by retention rate:

- + **High engaged:** 9-10/10 wks
- + **Medium engaged:** 5-8/10 wks
- + **Low engaged:** 1-4/10 wks

Mix of age, gender, occupation, and other DW behaviors (avg. stream count, skips, and minutes streamed)

Example: Discover Weekly

Sampling

Cohorts defined by retention rate:

- + **High engaged:** 9-10/10 wks
- + **Medium engaged:** 5-8/10 wks
- + **Low engaged:** 1-4/10 wks

Mix of age, gender, occupation, and other DW behaviors (avg. stream count, skips, and minutes streamed)

Content

Have a conversation:

- + Music preferences
- + Music habits
- + Attitudes toward Discover Weekly
- + Reasons for using Discover Weekly
- + Recent descriptions of Discover Weekly usage
- + Good experiences
- + Bad experiences
- + Walk through of current Discover Weekly
- + Behavior deep dive (list of interactions possible from DW page)
 - Have you done X? If yes, why? If no, why not?

Interviewing limitations

Generalizability

Small sample size from a single location. Findings should be considered hypotheses.

Demand effects

Participants may want to say what they think will make the interviewer happy.

May be un-naturalistic

In-lab interviews not where/how participants would normally be.

Interviewer bias

Interviewer goes in with preconceived notions, experiences, and background.

Interviewing limitations

Generalizability

Small sample size from a single location. Findings should be considered hypotheses.

Demand effects

Participants may want to say what they think will make the interviewer happy.

May be un-naturalistic

In-lab interviews not where/how participants would normally be.

Interviewer bias

Interviewer goes in with preconceived notions, experiences, and background.

Interviewing limitations

Generalizability

Small sample size from a single location. Findings should be considered hypotheses.

Demand effects

Participants may want to say what they think will make the interviewer happy.

May be un-naturalistic

In-lab interviews not where/how participants would normally be.

Interviewer bias

Interviewer goes in with preconceived notions, experiences, and background.

Interviewing limitations

Generalizability

Small sample size from a single location. Findings should be considered hypotheses.

Demand effects

Participants may want to say what they think will make the interviewer happy.

May be un-naturalistic

In-lab interviews not where/how participants would normally be.

Interviewer bias

Interviewer goes in with preconceived notions, experiences, and background.

Surveys

Why use a satisfaction survey?



Benchmark and track sentiment

Useful for seeing how sentiment moves over time.



Explore drivers of satisfaction

Potential drivers should be derived from qualitative analysis.



Create a labeled dataset

To explore what may proxies for user satisfaction.

To use as a target for prediction.

Why use a satisfaction survey?



Benchmark and track sentiment

Useful for seeing how sentiment moves over time.



Explore drivers of satisfaction

Potential drivers should be derived from qualitative analysis.



Create a labeled dataset

To explore what may proxies for user satisfaction.

To use as a target for prediction.

Why use a satisfaction survey?



Benchmark and track sentiment

Useful for seeing how sentiment moves over time.



Explore drivers of satisfaction

Potential drivers should be derived from qualitative analysis.



Create a labeled dataset

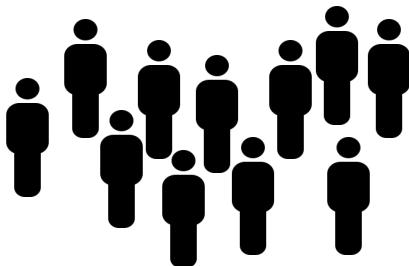
To explore what may proxies for user satisfaction.

To use as a target for prediction.

Survey sampling

Who do you want to
hear from?

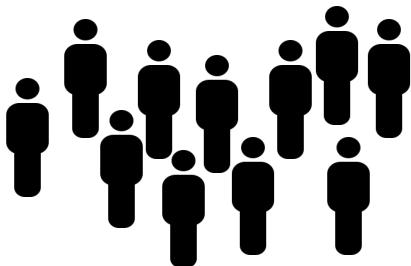
How do you define a user of
your product or feature?



Survey sampling

Who do you want to hear from?

How do you define a user of your product or feature?



Select a sampling design

Random: every member of population has equal chance to get selected

- + Better for general exploration
- + Easier analysis

Stratified: specific characteristics of individuals are represented in the sample and sample reflects true proportion of the population with those characteristics

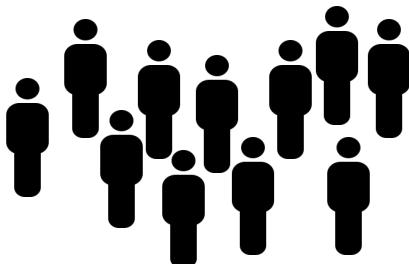
- + Good for exploring more specific hypothesis

Convenience: participants selected based on their convenience and availability

Survey sampling

Who do you want to hear from?

How do you define a user of your product or feature?



Select a sampling design

Random: every member of population has equal chance to get selected

- + Better for general exploration
- + Easier analysis

Stratified: specific characteristics of individuals are represented in the sample and sample reflects true proportion of the population with those characteristics

- + Good for exploring more specific hypothesis

Convenience: participants selected based on their convenience and availability

Determine a sample size

Power analysis! Covered in Hypothesis Testing.

Survey delivery



Email



Intercept



Survey delivery



Email



- + Don't need to worry about interrupting/ruining experience
- + Good for reflecting on a set of experiences



- + May be distant from the experience
- + May have lower response rate



Intercept

Survey delivery



Email



- + Don't need to worry about interrupting/ruining experience
- + Good for reflecting on a set of experiences



- + May be distant from the experience
- + May have lower response rate



Intercept

- + Capture response in context
- + Good for reflecting on a single experience with the system
- + May have higher response rate

- + Determining appropriate trigger time is challenging
- + Disrupts the user experience
- + Interrupts or alters behavior we are trying to track

Designing your satisfaction survey

Survey design should be based on qualitative findings!

Is it even reasonable to ask this question?



What were drivers of satisfaction?

What were people's goals?

What questions should you use from existing literature and scales?

Writing your questions - survey best practices

Debias your questions and responses.

Writing your questions - survey best practices

Debias your questions and responses.

How satisfied are you with the thing?



A horizontal scale consisting of ten light gray rectangular boxes arranged in a row. Each box contains a black number from 1 to 10, spaced evenly apart. The first box contains '1', the second '2', the third '3', the fourth '4', the fifth '5', the sixth '6', the seventh '7', the eighth '8', the ninth '9', and the tenth '10'.

Writing your questions - survey best practices

Debias your questions and responses.



How satisfied are you with the thing?

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

Writing your questions - survey best practices

Debias your questions and responses.



How satisfied are you with the thing?

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

Avoid implicit assumptions

Sample question assumes:

- + Respondent has used and remembers using the thing
- + No possibility for dissatisfaction

Writing your questions - survey best practices

Debias your questions and responses.



How satisfied are you with the thing?



Avoid implicit assumptions

Sample question assumes:

- + Respondent has used and remembers using the thing
- + No possibility for dissatisfaction

Scale should reflect the question

- + Scale shouldn't consist of abstract values
- + Scale should have words that are meaningful answers to the question

Writing your questions - survey best practices

Be like your respondents

Writing your questions - survey best practices

Be like your respondents

To what extent do you find that you achieve your goal(s) with our platform?

A horizontal scale consisting of ten light gray rectangular boxes, each containing a number from 1 to 10. The numbers are centered in their respective boxes.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

Writing your questions - survey best practices

Be like your respondents



To what extent do you find that you achieve your goal(s) with our platform?

1 2 3 4 5 6 7 8 9 10

Writing your questions - survey best practices

Be like your respondents



To what extent do you find that you achieve your goal(s) with our platform?

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

Questions and responses reflect their experiences

- + May not think in terms of goals or have multiple with different answers
- + Give them a way to tell you this doesn't make sense to them

Writing your questions - survey best practices

Be like your respondents



To what extent do you find that you achieve your goal(s) with our platform?

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

Questions and responses reflect their experiences

- + May not think in terms of goals or have multiple with different answers
- + Give them a way to tell you this doesn't make sense to them

Use words they would use

- + Words like "goal" or "platform" may mean something different to respondents than they do to you
- + Try to simplify question structure and use common language

Writing your questions - survey best practices

Be smart about open ends

Writing your questions - survey best practices

Be smart about open ends

Open ends require the highest cognitive (thinking) and physical (typing) load for your respondents, so don't use these willy nilly and make them optional.

Writing your questions - survey best practices

Be smart about open ends

Open ends require the highest cognitive (thinking) and physical (typing) load for your respondents, so don't use these willy nilly and make them optional.

- ★ When you need to know why (e.g., paired with a satisfaction rating)
- ★ When you can't possibly narrow into a set of response options (e.g., when you need to know the context)
- ★ When you don't know enough to narrow options (e.g., you didn't run qual first 😞 😠)

Usability testing your survey

What are we trying to learn?

- + Do people understand the questions?
- + How do they interpret and answer?
- + How do they use the scale?
- + What are we missing?

Usability testing your survey

What are we trying to learn?

- + Do people understand the questions?
- + How do they interpret and answer?
- + How do they use the scale?
- + What are we missing?

How do we learn it?

Recruit 5-10 people from the defined population

Create a setting as close to how they could experience the survey as possible.

Have them complete the survey while talking through everything they are thinking as they do. Ask meaningful probes (e.g., what made you respond that way, tell me more about your thinking).

Have them walk you through their experiences with the system they are evaluating to compare with their survey responses.

Ideally, iterate.

Example: Discover Weekly

We designed the survey to understand satisfaction on different levels, reflecting what we learned about satisfaction from the interviews.

OVERALL SATISFACTION

+ drivers:

- usability
- usefulness
- delight

USER GOAL RANK

- listen right now
- background music
- fits a specific activity
- save for later
- artist exploration
- genre exploration

THIS WEEK SATISFACTION

+ drivers:

- goal achievement
- transparency
- fit
- track love
- track annoyance

PERSONAL PREFERENCES FOR DISCOVER WEEKLY

- cohesion
- ‘ambitiousness’ of recommendations
- effort

Example: Discover Weekly

We designed the survey to understand satisfaction on different levels, reflecting what we learned about satisfaction from the interviews.

OVERALL SATISFACTION

+ drivers:

- usability
- usefulness
- delight

USER GOAL RANK

- listen right now
- background music
- fits a specific activity
- save for later
- artist exploration
- genre exploration

THIS WEEK SATISFACTION

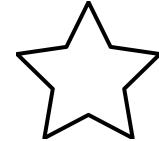
+ drivers:

- goal achievement
- transparency
- fit
- track love
- track annoyance

PERSONAL PREFERENCES FOR DISCOVER WEEKLY

- cohesion
- ‘ambitiousness’ of recommendations
- effort

Example: Discover Weekly



BEST PRACTICES

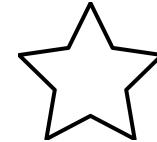
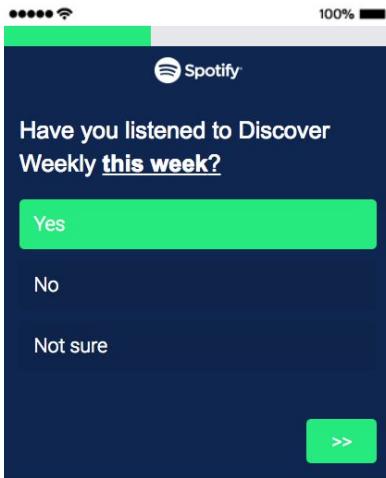


USABILITY



INTERVIEWS

Example: Discover Weekly



BEST PRACTICES

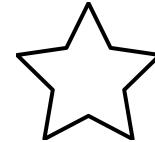


USABILITY



INTERVIEWS

Example: Discover Weekly



BEST PRACTICES



USABILITY

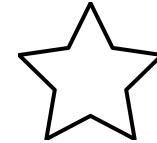


INTERVIEWS



No assumptions
about memory of
usage

Example: Discover Weekly



BEST PRACTICES



USABILITY



INTERVIEWS



No assumptions
about memory of
usage



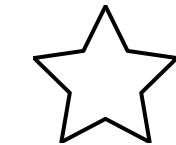
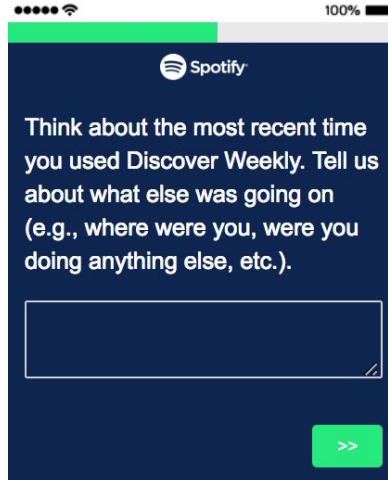
Needed to clearly
highlight time
frame of questions

Example: Discover Weekly



No assumptions
about memory of
usage

Needed to clearly
highlight time
frame of questions



BEST PRACTICES



USABILITY



INTERVIEWS

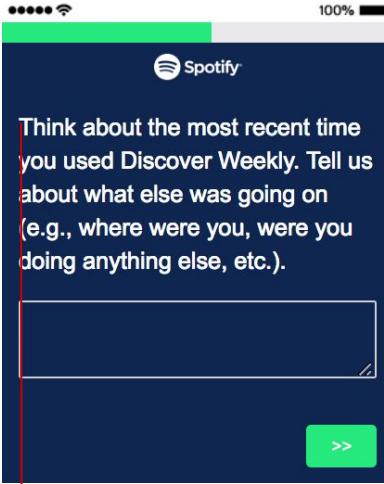
Example: Discover Weekly



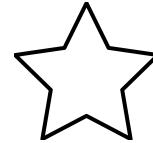
No assumptions
about memory of
usage



Needed to clearly
highlight time
frame of questions



Jog memory of
specific experience
to help situate
respondent in the
appropriate
context



BEST PRACTICES



USABILITY



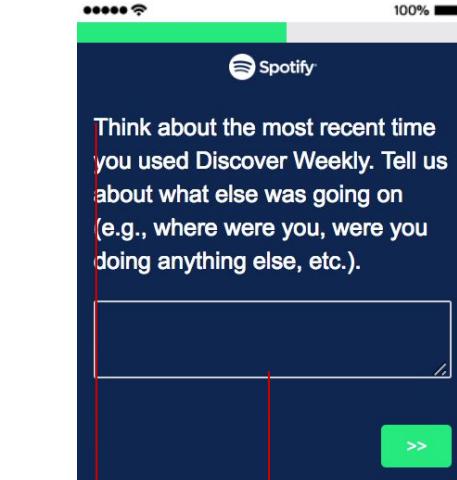
INTERVIEWS

Example: Discover Weekly

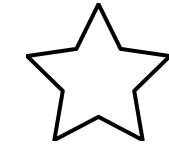


★
No assumptions
about memory of
usage

Needed to clearly
highlight time
frame of questions



★
Jog memory of
specific experience
to help situate
respondent in the
appropriate
context



BEST PRACTICES



USABILITY



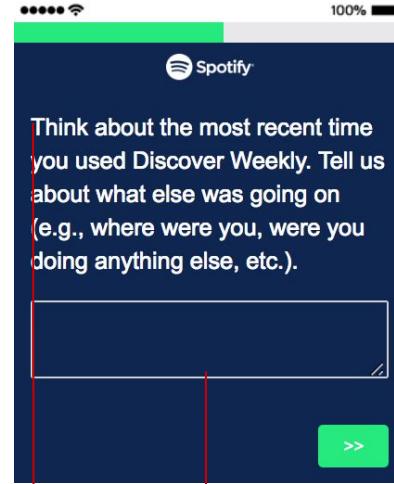
INTERVIEWS

Example: Discover Weekly

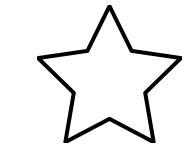


★
No assumptions
about memory of
usage

Needed to clearly
highlight time
frame of questions



★
Jog memory of
specific experience
to help situate
respondent in the
appropriate
context



BEST PRACTICES



USABILITY



INTERVIEWS



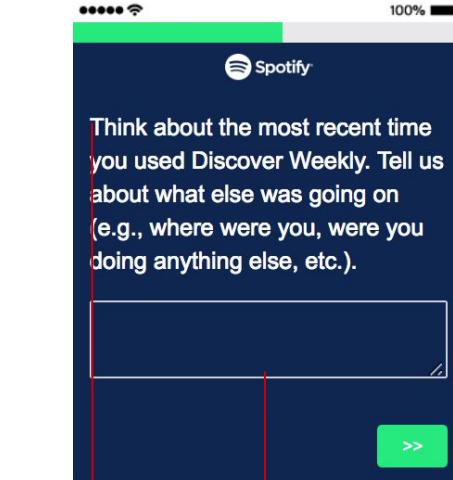
★
Open end because
we can't possibly
know the right
response options

Example: Discover Weekly



No assumptions
about memory of
usage

Needed to clearly
highlight time
frame of questions



Jog memory of
specific experience
to help situate
respondent in the
appropriate
context

Open end because
we can't possibly
know the right
response options



BEST PRACTICES



USABILITY



INTERVIEWS

Balanced wording

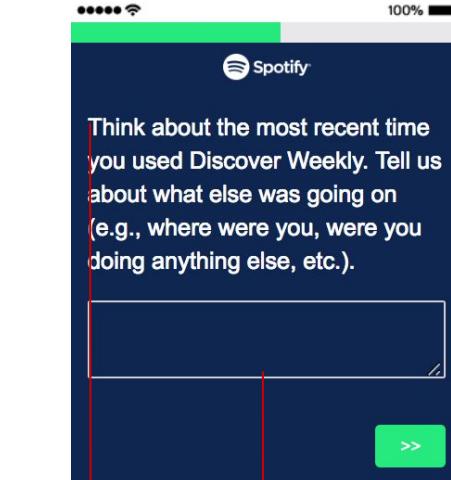


Example: Discover Weekly



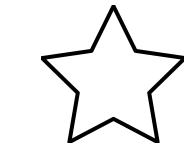
No assumptions
about memory of
usage

Needed to clearly
highlight time
frame of questions



Jog memory of
specific experience
to help situate
respondent in the
appropriate
context

Open end because
we can't possibly
know the right
response options



BEST PRACTICES



USABILITY



INTERVIEWS

Balanced wording



Scale adapted from
7-pt because
participants
couldn't distinguish
meaningfully
between options



Thinking specifically about this week's Discover Weekly, how satisfied or dissatisfied were you?

Very dissatisfied

Somewhat dissatisfied

Neither satisfied nor dissatisfied

Somewhat satisfied

Very satisfied

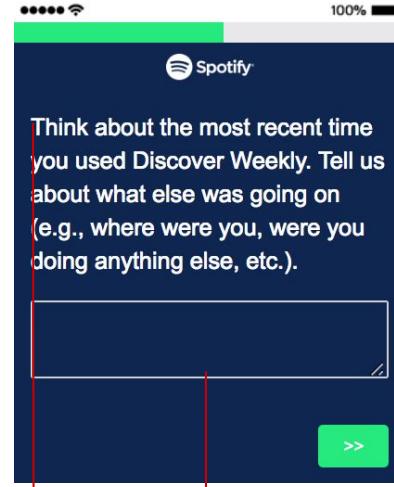
(Optional) Why did you give this answer?

Example: Discover Weekly



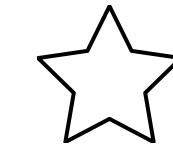
No assumptions
about memory of
usage

Needed to clearly
highlight time
frame of questions



Jog memory of
specific experience
to help situate
respondent in the
appropriate
context

Open end because
we can't possibly
know the right
response options



BEST PRACTICES



USABILITY



INTERVIEWS

Balanced wording



Scale adapted from
7-pt because
participants
couldn't distinguish
meaningfully
between options



Thinking specifically about this week's Discover Weekly, how satisfied or dissatisfied were you?

Very dissatisfied

Somewhat dissatisfied

Neither satisfied nor dissatisfied

Somewhat satisfied

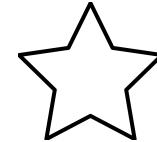
Very satisfied

Optional open-end
to provide more
insight into key
ratings



(Optional) Why did you give this answer?

Example: Discover Weekly



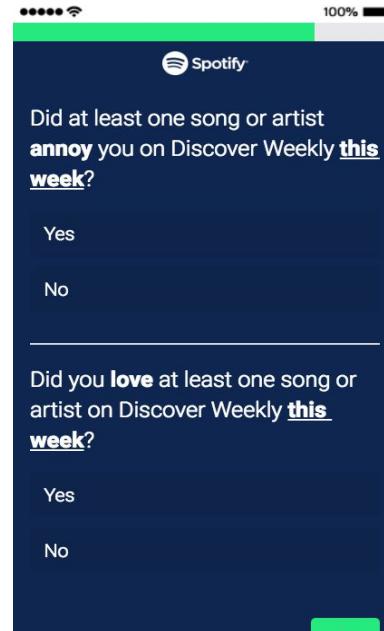
BEST PRACTICES



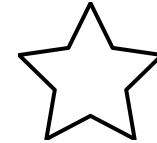
USABILITY



INTERVIEWS



Example: Discover Weekly



BEST PRACTICES



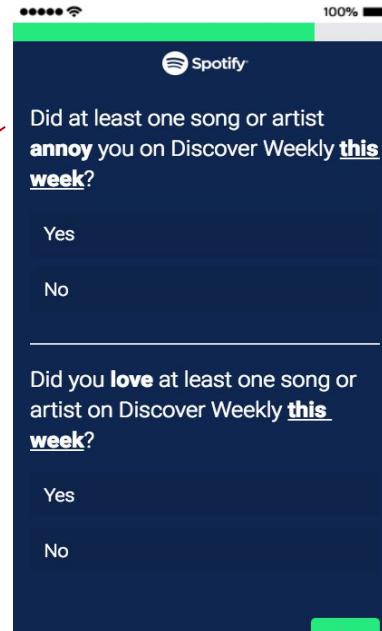
USABILITY



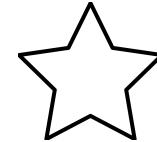
INTERVIEWS



Based on participant definitions of good and bad experiences with Discover Weekly



Example: Discover Weekly



BEST PRACTICES



USABILITY



INTERVIEWS

The screenshot shows a Spotify mobile application interface. At the top, it displays signal strength, battery level at 100%, and the Spotify logo. Below this is a survey section with two questions:

Did at least one song or artist annoy you on Discover Weekly this week?
Yes
No

Did you love at least one song or artist on Discover Weekly this week?
Yes
No

A red line connects a speech bubble icon on the left to the survey questions, indicating that the survey was based on participant definitions of good and bad experiences with Discover Weekly. Another red line connects a smartphone icon on the right to the survey results, with the accompanying text explaining that scales were adjusted to binary because participants felt it was a simpler way to capture their experience.

Based on participant definitions of good and bad experiences with Discover Weekly

Scales adjusted to binary because participants felt it was a simpler way to capture their experience

Survey limitations

Unnatural

Responding to questions that are qualitative in nature is difficult for people to quantify, as it assumes their experiences can be reduced to a single, evenly spaced dimension.

Response bias

Despite sampling efforts, the people who decide to respond to your survey are likely to be systematically different from the people who decide not to.

Individual differences in scale use

Some people have quirks with how they use scales. For example, some refuse to ever use the top of the scale believing that there's always room for improvement, while others avoid the bottom half of the scale, not wanting to be too negative.

Survey limitations

Unnatural

Responding to questions that are qualitative in nature is difficult for people to quantify, as it assumes their experiences can be reduced to a single, evenly spaced dimension.

Response bias

Despite sampling efforts, the people who decide to respond to your survey are likely to be systematically different from the people who decide not to.

Individual differences in scale use

Some people have quirks with how they use scales. For example, some refuse to ever use the top of the scale believing that there's always room for improvement, while others avoid the bottom half of the scale, not wanting to be too negative.

Survey limitations

Unnatural

Responding to questions that are qualitative in nature is difficult for people to quantify, as it assumes their experiences can be reduced to a single, evenly spaced dimension.

Response bias

Despite sampling efforts, the people who decide to respond to your survey are likely to be systematically different from the people who decide not to.

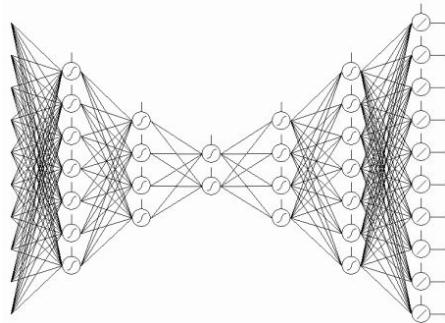
Individual differences in scale use

Some people have quirks with how they use scales. For example, some refuse to ever use the top of the scale believing that there's always room for improvement, while others avoid the bottom half of the scale, not wanting to be too negative.

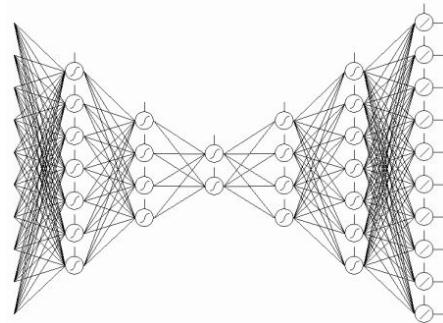
Quantitative Data Collection



1
user begins
session



4
user responds to
system decision



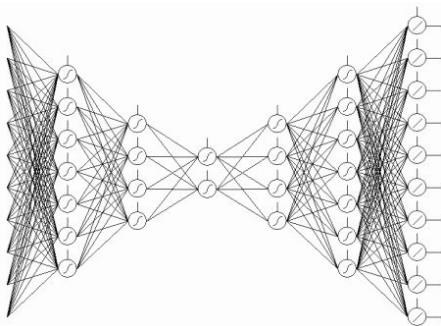
2
system observes
user state

3
system generates
decision

5
system observes
response

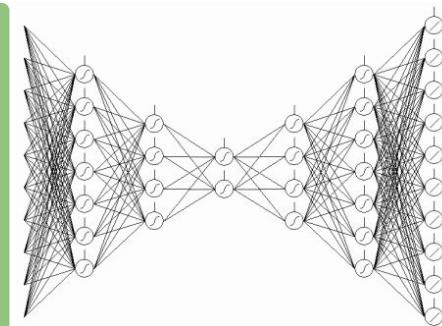
6
system interprets
response

**pre-presentation
features**



**recommendation
system**

**post-presentation
features**



reward prediction

Quantitative Data Collection

objective: collect user behavior data affected to the quality of a system decisions.

system decision: recommendation, search results, search session, interface.

quality: item relevance, search satisfaction, entertainment.

user behavior data: logged interactions attributable to system decisions.



quantitative behavioral signals should be developed
and vetted in collaboration with qualitative research

Explicit Feedback

explicit feedback: directly ask the user for feedback on a system decisions

- *examples*
 - stars
 - ratings
 - reactions
- *pros*
 - user control
- *cons:*
 - inconsistent user intent (i.e. what does X mean?)
 - sparse
 - invasive

+ Monolith - Remastered T. Rex Electric Warrior [Expa... 6 days ago

+ Sally Go Round The Roses The Jaynetts Sally Go 'Round The ... 6 days ago

(+) + Wichita Lineman The Meters Struttin' 6 days ago   ...

+ Ad Gloriam Le Orme Ad Gloriam 6 days ago Like

+ Lonely Alan Vega Untitled 6 days ago

Implicit Feedback

implicit feedback: observe the user for feedback on a system decisions

- *examples*
 - clicks
 - streams
 - abandonment
- *pros*
 - denser
 - unobtrusive
- *cons:*
 - inconsistent user intent (i.e. what does X mean?)
 - no user control

- | | | | | | |
|---|--------------------------|--------------|----------------------------|------------|---|
| + | Magic in the Moonlight | Magic Tramps | Kickin' Up Moonlight ... | 6 days ago | |
| + | Atlantis | Dory Previn | Dory Previn | 6 days ago | |
| + | Monolith - Remastered | T. Rex | Electric Warrior [Expa...] | 6 days ago |    |
| + | Sally Go Round The Roses | The Jaynetts | Sally Go 'Round The ... | 6 days ago | |

Common Behaviors

| behavior | question | example |
|---------------------|--|--|
| attention | did the user see the system decision? | <i>page load</i> |
| interaction | did the user interact with the recommendation? | <i>track stream</i> |
| task success | was the user satisfied by the recommendation? | <i>completed playlist</i> |
| retention | is this a useful tool for the user? | <i>return to recommendation system</i> |

Common Behaviors

| behavior | question | example |
|---------------------|--|--|
| attention | did the user see the system decision? | <i>page load</i> |
| interaction | did the user interact with the recommendation? | <i>track stream</i> |
| task success | was the user satisfied by the recommendation? | <i>completed playlist</i> |
| retention | is this a useful tool for the user? | <i>return to recommendation system</i> |

attention

- did the user see the system decision?
- importance
 - cannot attribute interaction to anything unattended
 - should attribute negative feedback to anything attended and not interacted with

attention: page load

Browse

OVERVIEW PODCASTS CHARTS GENRES & MOODS NEW RELEASES **DISCOVER** CONCERTS VIDEOS

Playlists made just for you

 Discover Weekly
Your weekly mixtape of fresh music. Enjoy new discoveries and deep cuts chosen just for you...
MADE FOR FERNANDO DIAZ

 Your Release Radar
Never miss a new release! Catch all the latest music from artists you care about, plus new singl...
MADE FOR FERNANDO DIAZ

Top recommendations for you



MADE FOR FERNANDO

Discover Weekly

Your weekly mixtape of fresh music. Enjoy new discoveries and deep cuts chosen just for you. Updated every Monday, so save your favorites!

Made for Fernando Diaz by Spotify - 30 songs, 1 hr 48 min

PLAY **FOLLOWING** **...**

Q Filter

| TITLE | ARTIST | ALBUM |
|--------|--------------|--------------------------|
| + Time | Richard Hell | Spurts: The Richard H... |
| | | 6 days ago |

attention: page scroll



The image shows a Spotify interface for a 'Discover Weekly' playlist. At the top, it says 'MADE FOR FERNANDO' and 'Discover Weekly'. Below that, it says 'Your weekly mixtape of fresh music. Enjoy new discoveries and deep cuts chosen just for you. Updated every Monday, so save your favorites!' and 'Made for Fernando Diaz by Spotify - 30 songs, 1 hr 48 min'. There are buttons for 'PLAY', 'FOLLOWING', and three dots. On the right, it shows 'FOLLOWER 1' and a 'Download' button with a toggle switch. Below the header is a search bar with 'Filter'. The main content is a table with columns for 'TITLE', 'ARTIST', and 'ALBUM'. The table lists ten songs:

| TITLE | ARTIST | ALBUM | LAST LISTENED |
|--------------------------|--------------|---------------------------|---------------|
| Time | Richard Hell | Spurts: The Richard H... | 6 days ago |
| Forever | Pete Drake | Forever - Single | 6 days ago |
| Magic in the Moonlight | Magic Tramps | Kickin' Up Moonlight ... | 6 days ago |
| Atlantis | Dory Previn | Dory Previn | 6 days ago |
| Monolith - Remastered | T. Rex | Electric Warrior [Expa... | 6 days ago |
| Sally Go Round The Roses | The Jaynetts | Sally Go 'Round The ... | 6 days ago |
| Wichita Lineman | The Meters | Struttin' | 6 days ago |
| Ad Gloriam | Le Orme | Ad Gloriam | 6 days ago |
| Lonely | Alan Vega | Untitled | 6 days ago |

attention: cursor-tracking

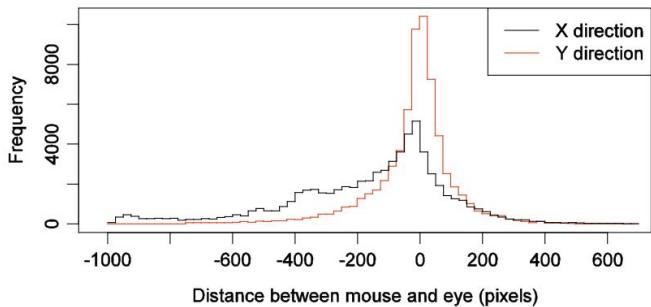
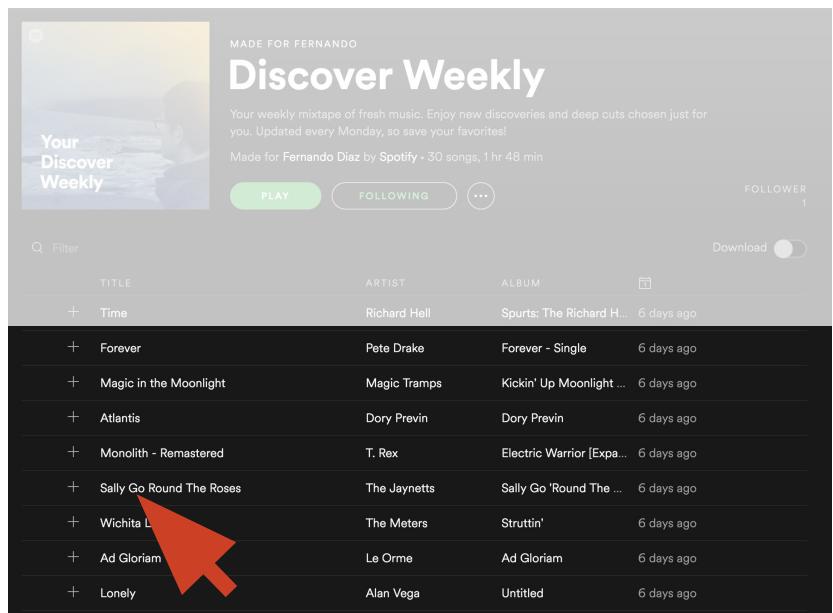


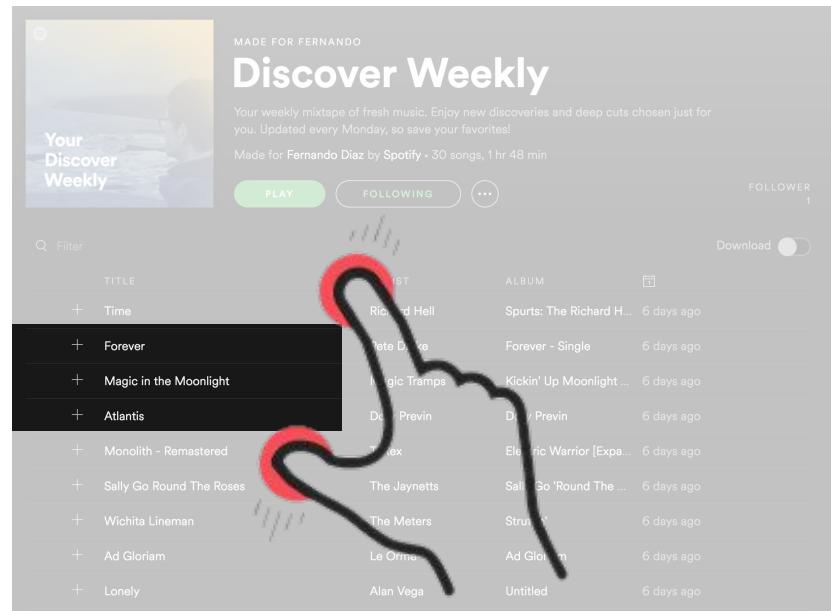
Figure 2: Histogram outlines of distance from mouse to eye, broken out separately for the X and Y directions. Each step in the histogram represents a bucket of 25 pixels. A negative X distance means that the eye was to the left of the mouse, and a negative Y distance means that the eye was below the mouse.



attention: touch gestures



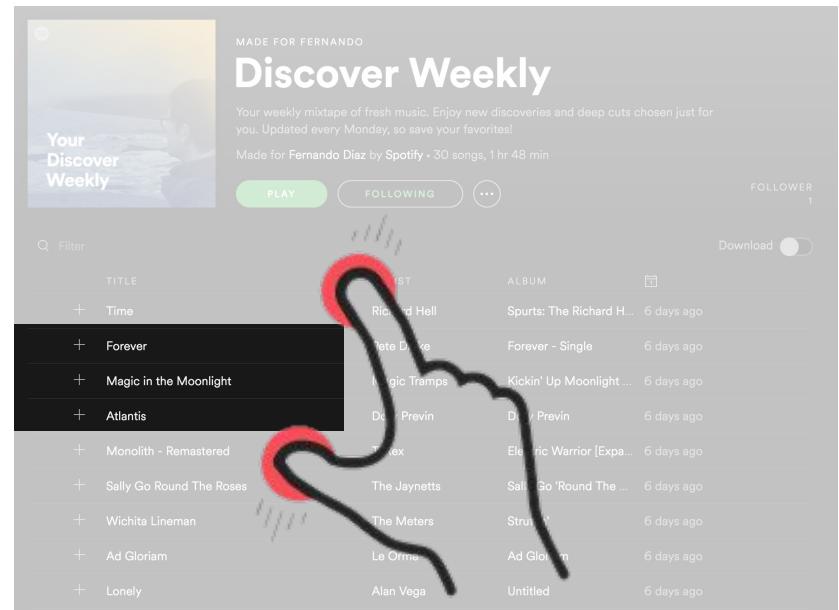
Figure 2. An illustrative example of a heatmap on a Web page generated from mobile device users' viewport data.



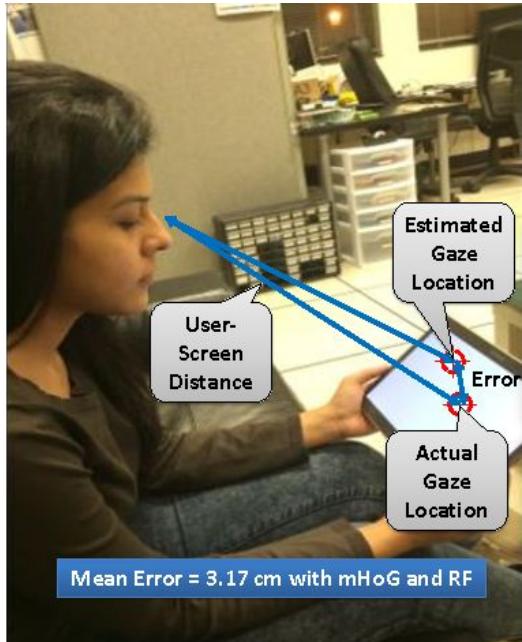
J. Huang and A. M. Diriye. Web user interaction mining from touch-enabled mobile devices. In Proceedings of the sixth symposium on human-computer interaction and information retrieval (HCIR 2012), 2012.

attention: touch gestures

| Group | Feature Description |
|-------------|---|
| Dwell | $dwell$: time of the page view in seconds $gestcnt$: number of touch gestures $gestfreq$: gestent/dwell $pressure$: average touch pressure, normalized to the range of 0 and 1 $touchsize$: average touch size, normalized to the range of 0 and 1 |
| Zooming | $zoomcnt$: number of zooming gestures $zoomfreq$: zoomcnt / dwell $zoomdist$: accumulated zooming scale changes $zoomspeed$: zoomdist / dwell $zoommax$: maximum zooming scale |
| Swiping | $swipecnt$: number of vertical swipes $swipefreq$: swipecnt / dwell $swipedist$: total vertical swipe distance $swipespeed$: swipedist / dwell $swipemax$: maximum vertical swipe coordinate |
| Inactivity | $inactive_total$: accumulated inactive time $inactive_pct$: inactive_total / dwell |
| Transitions | $inactive_avg$: average inactive time $inactive_max$: maximum inactive time $s_i-s_j_cnt$: number of transitions from state s_i to observation s_j during the page view $transitions_cnt$: total number of observation transitions during the page view $s_i-s_j_prob$: $s_i-s_j_cnt$ / $transitions_cnt$ |
| User | $user_id$: the ID of the user who viewed the page |



attention: eye-tracking



The image shows a Spotify 'Discover Weekly' playlist page. At the top, it says 'MADE FOR FERNANDO' and 'Discover Weekly'. Below that, it says 'Your weekly mixtape of fresh music. Enjoy new discoveries and deep cuts chosen just for you. Updated every Monday, so save your favorites!' and 'Made for Fernando Diaz by Spotify - 30 songs, 1 hr 48 min'. There are buttons for 'PLAY', 'FOLLOWING', and '...'. On the right, it says 'FOLLOWER 1' and has a 'Download' button. The main part of the image is a table listing 30 songs in the playlist:

| TITLE | ARTIST | ALBUM | LAST LISTENED |
|--------------------------|--------------|---------------------------|---------------|
| Time | Richard Hell | Spurts: The Richard H... | 6 days ago |
| Forever | Pete Drake | Forever - Single | 6 days ago |
| Magic in the Moonlight | Magic Tramps | Kickin' Up Moonlight ... | 6 days ago |
| Atlantis | Dory Previn | Dory Previn | 6 days ago |
| Monolith - Remastered | T. Rex | Electric Warrior [Expa... | 6 days ago |
| Sally Go Round The Roses | The Jaynetts | Sally Go 'Round The ... | 6 days ago |
| Wichita Lineman | The Meters | Struttin' | 6 days ago |
| Ad Gloriam | Le Orme | Ad Gloriam | 6 days ago |
| Lonely | Alan Vega | Untitled | 6 days ago |

Qiong Huang, Ashok Veeraraghavan, and Ashutosh Sabharwal. Tabletgaze: A dataset and baseline algorithms for unconstrained appearance-based gaze estimation in mobile tablets. CoRR, abs/1508.012442015, 2015.

Common Behaviors

| behavior | question | example |
|---------------------|--|--|
| attention | did the user see the system decision? | <i>page load</i> |
| interaction | did the user interact with the recommendation? | <i>track stream</i> |
| task success | was the user satisfied by the recommendation? | <i>completed playlist</i> |
| retention | is this a useful tool for the user? | <i>return to recommendation system</i> |

interaction

- did the user interact with the recommendation?
- importance
 - most recommendations are experiential goods

Common Interaction Behaviors

| behavior | segment | object | class |
|------------------|-------------------|----------------------------|-----------|
| examine | view, listen | select | |
| retain | print | bookmark, save, delete | subscribe |
| reference | copy/paste, quote | forward, reply, link, cite | |
| annotate | mark up | rate, publish | organize |
| create | type, edit | author | |

D. Oard and J. Kim. Modeling information content using observable behavior. In Proceedings of the ASIST annual meeting, 2001.
D. Kelly and J. Teevan. 2003. Implicit feedback for inferring user preference: a bibliography. SIGIR Forum 37, 2 (September 2003).

interaction: clicks

Montreal - Wikipedia
<https://en.wikipedia.org/wiki/Montreal> ▾
Montreal is the most populous municipality in the Canadian province of Quebec and the second-most populous municipality in Canada. Originally called ...

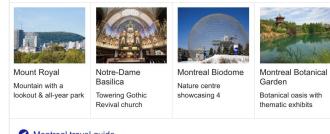
Old Montreal · Demographics of Montreal · History of Montreal · Island of Montreal

Tourisme Montréal: Discover your Montréal

<https://www.mtl.org/en> ▾

Find all the information you need for your trip to Montréal, from transportation and hotels to activities and restaurants.

Things to do in Montréal



Montréal travel guide

Visit the Ville de Montréal website - Ville de Montréal - Official city portal

ville.montreal.qc.ca/portal/page?_pageid=5977,804&_dad=portal... ▾
Community corner: Info-TravauxFind out about current roadwork and make travelling easier (in French). Follow us on Twitter - Participez aux Grandes conversations ...

Top stories

Montreal Canadiens place Kenny Agostino and Michael Chaput on waivers

Eyes On The Prize · 1 hour ago

Canadiens @ Senators game recap: Ottawa 3-0 in final pre-season match

Habs Eyes on the Prize · 8 hours ago



Montreal

Municipality in Québec

Montreal is the largest city in Canada's Quebec province. It's set on an island in the Saint Lawrence River and named after Mt. Royal, the triple-peaked hill at its heart. Its boroughs, many of which were once independent towns, range from the modern skyscrapers rising from downtown, French colonial Vieux-Montréal — with the Gothic Revival Notre-Dame Basilica at its centre — to bohemian Plateau.

Area: 431.5 km²

Weather: 11 °C Wind SW at 10 kmh, 62% Humidity

Local time: Sunday 2:08 p.m.

Population: 1.741 million (2014) United Nations

Plan a trip

Montréal travel guide

3-star hotel averaging \$155, 5-star averaging \$491

Upcoming events

Colleges and Universities: McGill University, MORE

Neighborhoods: Downtown Montreal, Old Montreal, MORE

Did you know: Montréal is the largest population centre in Quebec by population (1,407,963), wikipedia.org

People also search for



[View 15+ more](#)

[Feedback](#)

| | | | | | |
|--|---|--------------------------|--------------|----------------------------|------------|
| | + | Monolith - Remastered | T. Rex | Electric Warrior [Expo...] | 6 days ago |
| | + | Sally Go Round The Roses | The Jaynetts | Sally Go 'Round The ... | 6 days ago |
| | + | Wichita Lineman | The Meters | Struttin' | 6 days ago |
| | + | Ad Gloriam | Le Orme | Ad Gloriam | 6 days ago |
| | + | Lonely | Alan Vega | Untitled | 6 days ago |

Thorsten Joachims. Optimizing search engines using clickthrough data. In Kdd '02: proceedings of the eighth acm sigkdd international conference on knowledge discovery and data mining, 133--142, 2002.

interaction: consumption time

Montreal - Wikipedia

<https://en.wikipedia.org/wiki/Montreal> ▾
Montreal is the most populous municipality in the Canadian province of Quebec and the second-most populous municipality in Canada. Originally called ...
Old Montreal · Demographics of Montreal · History of Montreal · Island of Montreal

Tourisme Montréal: Discover your Montréal

<https://www.mtl.org/en> ▾
Find all the information you need for your trip to Montréal, from transportation and hotels to activities and restaurants.

Things to do in Montreal



Montréal travel guide

Visit the Ville de Montréal website - Ville de Montréal - Official city portal
ville.montreal.qc.ca/portal/page?_pageid=5977,804&_dad=portal... ▾
Community corner: Info-TravauxFind out about current roadwork and make travelling easier (in French). Follow us on Twitter - Participez aux Grandes conversations ...

Top stories

Montreal Canadiens place Kenny Agostino and Michael Chaput on waivers



Eyes On The Prize | 1 hour ago

Canadiens @ Senators game recap: Montreal falls 3-0 in final pre-season match

Habs Eyes on the Prize | 8 hours ago



Montreal

Municipality in Québec

Montreal is the largest city in Canada's Quebec province. It's set on an island in the Saint Lawrence River and named after Mt. Royal, the triple-peaked hill at its heart. Its boroughs, many of which were once independent towns, range from the ultramodern to the traditional, French colonial Vieux-Montréal — with the Gothic Revival Notre-Dame Basilica at its centre — to bohemian Plateau.

Area: 431.5 km²

Weather: 11 °C, Wind SW at 10 kmh, 62% Humidity

Local time: Sunday 2:08 p.m.

Population: 1.741 million (2014) United Nations

Plan a trip

Montréal travel guide

3-star hotel averaging \$155, 5-star averaging \$491

Upcoming events

Colleges and Universities: McGill University, MORE

Neighborhoods: Downtown Montreal, Old Montreal, MORE

Did you know: Montréal is the largest population centre in Quebec by population (1,407,963), wikipedia.org

People also search for



[View 15+ more](#)



| | | | |
|--|--------------|----------------------------|------------|
| + Monolith - Remastered | T. Rex | Electric Warrior [Expo...] | 6 days ago |
| + Sally Go Round The Roses | The Jaynetts | Sally Go 'Round The ... | 6 days ago |
| ▶ + Wichita Lineman | The Meters | Struttin' | 6 days ago |
| + Ad Gloriam | Le Orme | Ad Gloriam | 6 days ago |
| + Lonely | Alan Vega | Untitled | 6 days ago |

M. Morita and Y. Shinoda. Information filtering based on user behavior analysis and best match text retrieval. SIGIR 1994.
M. Claypool, P. Le, M. Wased, D. Brown. Implicit interest indicators. In Proceedings of the 6th international conference on intelligent user interfaces, IUI 2001.
Kim, Hassan, White, Zitouni. "Modeling dwell time to predict click-level satisfaction", WSDM 2014.

interaction: saving

Montreal - Wikipedia

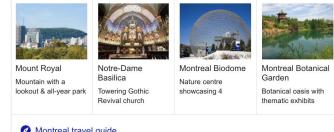
<https://en.wikipedia.org/wiki/Montreal> ▾
Montreal is the most populous municipality in the Canadian province of Quebec and the second-most populous municipality in Canada. Originally called ...
Old Montreal · Demographics of Montreal · History of Montreal · Island of Montreal

Tourisme Montréal: Discover your Montréal

<https://www.mtl.org/en> ▾

Find all the information you need for your trip to Montréal, from transportation and hotels to activities and restaurants.

Things to do in Montreal



Montréal travel guide

Visit the Ville de Montréal website - Ville de Montréal - Official city portal
ville.montreal.qc.ca/portal/page?_pageid=5977,80481579&_dad=portal... ▾
Community corner: Info-TravauxFind out about current roadwork and make travelling easier (in French). Follow us on Twitter - Participez aux Grandes conversations ...

Top stories



Montreal

Municipality in Québec

Montreal is the largest city in Canada's Québec province. It's set on an island in the Saint Lawrence River and named after Mt. Royal, the triple-peaked hill at its heart. Its boroughs, many of which were once independent towns, range from the modern skyscrapers rising from the old port to the Gothic Revival Notre-Dame Basilica at its centre – to bohemian Plateau.

Area: 431.5 km²

Weather: 11 °C, Wind SW at 10 kmh, 62% Humidity

Local time: Sunday 2:08 p.m.

Population: 1.741 million (2014) United Nations

Plan a trip

Montréal travel guide

3-star hotel averaging \$156, 5-star averaging \$491

Upcoming events

Colleges and Universities: McGill University, MORE

Neighborhoods: Downtown Montreal, Old Montreal, MORE

Did you know: Montréal is the largest population centre in Québec by population (1,407,963). [wikipedia.org](#)

People also search for



Feedback



| | | | |
|--|--------------|----------------------------|------------|
| + Monolith - Remastered | T. Rex | Electric Warrior [Expo...] | 6 days ago |
| + Sally Go Round The Roses | The Jaynetts | Sally Go 'Round The ... | 6 days ago |
| + Wichita Lineman | The Meters | Struttin' | 6 days ago |
| + Ad Gloriam | Le Orme | Ad Gloriam | 6 days ago |
| + Lonely | Alan Vega | Untitled | 6 days ago |

interaction: sharing

Montreal - Wikipedia

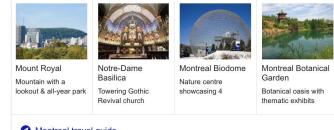
<https://en.wikipedia.org/wiki/Montreal> ▾
Montreal is the most populous municipality in the Canadian province of Quebec and the second-most populous municipality in Canada. Originally called ...
Old Montreal · Demographics of Montreal · History of Montreal · Island of Montreal

Tourisme Montréal: Discover your Montréal

<https://www.mtl.org/en> ▾

Find all the information you need for your trip to Montréal, from transportation and hotels to activities and restaurants.

Things to do in Montreal



Montréal travel guide

Visit the Ville de Montréal website - Ville de Montréal - Official city portal
ville.montreal.qc.ca/portal/page?_pageid=5977,804&_dad=portal... ▾
Community corner: Info-TravauxFind out about current roadwork and make travelling easier (in French). Follow us on Twitter - Participez aux Grandes conversations ...

Top stories

Montreal Canadiens place Kenny Agostino and Michael Chaput on waivers
[Eyes On The Prize](#) · 1 hour ago

Canadiens @ Senators game recap: Montreal falls 3-0 in final pre-season match
[Habs Eyes on the Prize](#) · 8 hours ago



Montreal

Municipality in Québec

Montreal is the largest city in Canada's Quebec province. It's set on an island in the Saint Lawrence River and named after Mt. Royal, the triple-peaked hill at its heart. Its boroughs, many of which were once independent towns, range from the modern skyscrapers rising from the old port to the preserved, French colonial Vieux-Montréal — with the Gothic Revival Notre-Dame Basilica at its centre — to bohemian Plateau.

Area: 431.5 km²

Weather: 11 °C, Wind SW at 10 kmh, 62% Humidity

Local time: Sunday 2:08 p.m.

Population: 1.741 million (2014) United Nations

Plan a trip

Montréal travel guide

3-star hotel averaging \$155, 5-star averaging \$491

Upcoming events

Colleges and Universities: McGill University, MORE

Neighborhoods: Downtown Montreal, Old Montreal, MORE

Did you know: Montréal is the largest population centre in Quebec by population (1,407,963). [wikipedia.org](#)

People also search for



[View 15+ more](#)

[Feedback](#)



| | | | |
|--|--------------|----------------------------|------------|
| + Monolith - Remastered | T. Rex | Electric Warrior [Expo...] | 6 days ago |
| + Sally Go Round The Roses | The Jaynetts | Sally Go 'Round The ... | 6 days ago |
| ▶ + Wichita Lineman | The Meters | Struttin' | 6 days ago |
| + Ad Gloriam | Le Orme | Ad Gloriam | 6 days ago |
| + Lonely | Alan Vega | Untitled | 6 days ago |

Anlei Dong, Ruiqiang Zhang, Pranam Kolari, Jing Bai, Fernando Diaz, Yi Chang, Zhaohui Zheng, and Hongyuan Zha. Time is of the essence: improving recency ranking using twitter data. In Wwww '10: proceedings of the 19th international conference on world wide web, 331--340, New York, NY, USA, 2010.

interaction: re-consumption

| | | | |
|----------------------------|--------------|---------------------------|--|
| + Monolith - Remastered | T. Rex | Electric Warrior [Expa... | 6 days ago |
| + Sally Go Round The Roses | The Jaynetts | Sally Go 'Round The ... | 6 days ago |
| ▶ + Wichita Lineman | The Meters | Struttin' | 6 days ago    ... |
| + Ad Glioriam | Le Orme | Ad Glioriam | 6 days ago  |
| + Lonely | Alan Vega | Untitled | 6 days ago |



| | | | |
|----------------------------|--------------|---------------------------|--|
| + Monolith - Remastered | T. Rex | Electric Warrior [Expa... | 6 days ago |
| + Sally Go Round The Roses | The Jaynetts | Sally Go 'Round The ... | 6 days ago |
| ▶ + Wichita Lineman | The Meters | Struttin' | 6 days ago    ... |
| + Ad Glioriam | Le Orme | Ad Glioriam | 6 days ago  |
| + Lonely | Alan Vega | Untitled | 6 days ago |



| | | | |
|----------------------------|--------------|---------------------------|--|
| + Monolith - Remastered | T. Rex | Electric Warrior [Expa... | 6 days ago |
| + Sally Go Round The Roses | The Jaynetts | Sally Go 'Round The ... | 6 days ago |
| ▶ + Wichita Lineman | The Meters | Struttin' | 6 days ago    ... |
| + Ad Glioriam | Le Orme | Ad Glioriam | 6 days ago  |
| + Lonely | Alan Vega | Untitled | 6 days ago |



interaction: knowledge

| | | | |
|----------------------------|--------------|-----------------------------|------------|
| + Monolith - Remastered | T. Rex | Electric Warrior [Expanded] | 6 days ago |
| + Sally Go Round The Roses | The Jaynetts | Sally Go Round The ... | 6 days ago |
| ▶ + Wichita Lineman | The Meters | Struttin' | 6 days ago |
| + Ad Gloriam | Le Orme | Ad Gloriam | 6 days ago |
| + Lonely | Alan Vega | Untitled | 6 days ago |



Common Behaviors

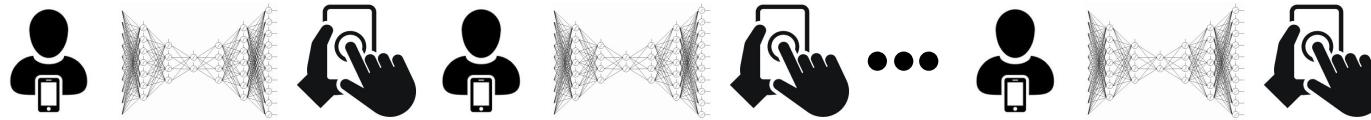
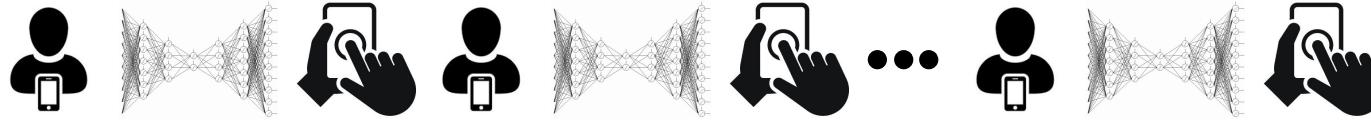
| behavior | question | example |
|---------------------|--|--|
| attention | did the user see the system decision? | <i>page load</i> |
| interaction | did the user interact with the recommendation? | <i>track stream</i> |
| task success | was the user satisfied by the recommendation? | <i>completed playlist</i> |
| retention | is this a useful tool for the user? | <i>return to recommendation system</i> |

Task-based Information Retrieval

A task is what someone does in order to achieve a goal (Hackos & Redish 1998). The goal of information searching is to find information, which supports task performance when an actor has insufficient knowledge. Thus, the search for information is a subtask in task performance.

Higher Level Intent

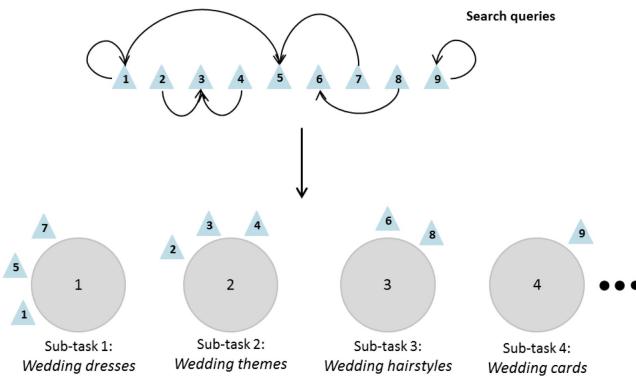
"A request is no more than a guess about the attributes a desired document is expected to have. In general, the response of the system to the request is then used to correct the initial guess for another try...It follows that retrieval systems should be designed to achieve maximum flexibility and rapid response in the trial-and-error process itself."



Observing Tasks

| QUERY and TIMESTAMP | GOAL # | MISSION # | DESCRIPTION |
|--|--------|-----------|---|
| hiking; san francisco Tue Apr 17 23:43:17 2007 (4m 17s) | 1 | 1 | MISSION 1: Find info on hiking opportunities in and around San Francisco |
| hiking; san francisco bay area Tue Apr 17 23:47:34 2007 (4m 59s) | 1 | 1 | GOAL 1: Find info on hiking trails in San Francisco and the Bay Area |
| ano nuevo state reserve Tue Apr 17 23:52:33 2007 (7m 54s) | 2 | 1 | GOAL 2: Navigate to Ano Nuevo State Reserve and find out about distances |
| ano nuevo state reserve; miles Wed Apr 18 00:00:27 2007 (3m 34s) | 2 | 1 | |
| nature trails; san francisco Wed Apr 18 00:04:01 2007 (16m 15s) | 1 | 1 | |
| lobos creek trail Wed Apr 18 00:20:16 2007 (0m 3s) | 3 | 1 | GOAL 3: Navigate to Lobos Creek Trail |
| china camp state park; san rafael Wed Apr 18 00:20:19 2007 (2m 35s) | 4 | 1 | GOAL 4: Navigate to China Camp, San Rafael and find out about distances |
| china camp; miles Wed Apr 18 00:22:54 2007 (20m 2s) | 4 | 1 | |
| hike; san francisco Wed Apr 18 00:42:56 2007 (3m 19s) | 1 | 1 | |
| fort funston Wed Apr 18 00:46:15 2007 (1h 51m 26s) | 5 | 1 | GOAL 5: Navigate to Fort Funston |
| | | | MISSION 2: Find info on car maintenance and repair |
| brake pads Wed Apr 18 03:36:47 2007 (16m 36s) | 6 | 2 | GOAL 6: Find info on brake pads |
| auto repair Wed Apr 18 03:53:23 2007 (8m 0s) | 7 | 2 | GOAL 7: Find info on an auto body shop in San Francisco |
| auto body shop Wed Apr 18 04:01:23 2007 (3m 31s) | 7 | 2 | |
| batteries Wed Apr 18 04:04:54 2007 (0m 29s) | 8 | 2 | |
| car batteries Wed Apr 18 04:05:23 2007 (2m 8s) | 8 | 2 | GOAL 8: Find info on purchasing a car battery |
| auto body shop; san francisco Wed Apr 18 04:07:31 2007 (3m 33s) | 7 | 2 | |
| buy car battery online free shipping Wed Apr 18 04:11:04 2007 | 8 | 2 | |

Modeling Tasks



| sub-task 1 | sub-task 2 | sub-task 3 |
|--------------------------|--------------------------|------------------------|
| wedding hairstyles | used wedding dresses | wedding card holders |
| wedding hair dos | colorful bridal gowns | indian wedding program |
| curly wedding hairstyles | preowned wedding dresses | wedding program |
| pictures of wedding hair | wedding attire | regency wedding cards |

R. Mehrotra, P. Bhattacharya, E. Yilmaz, "Deconstructing Complex Search Tasks: a Bayesian Nonparametric Approach for Extracting Sub-tasks", Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 599–605.

Predicting Task Success

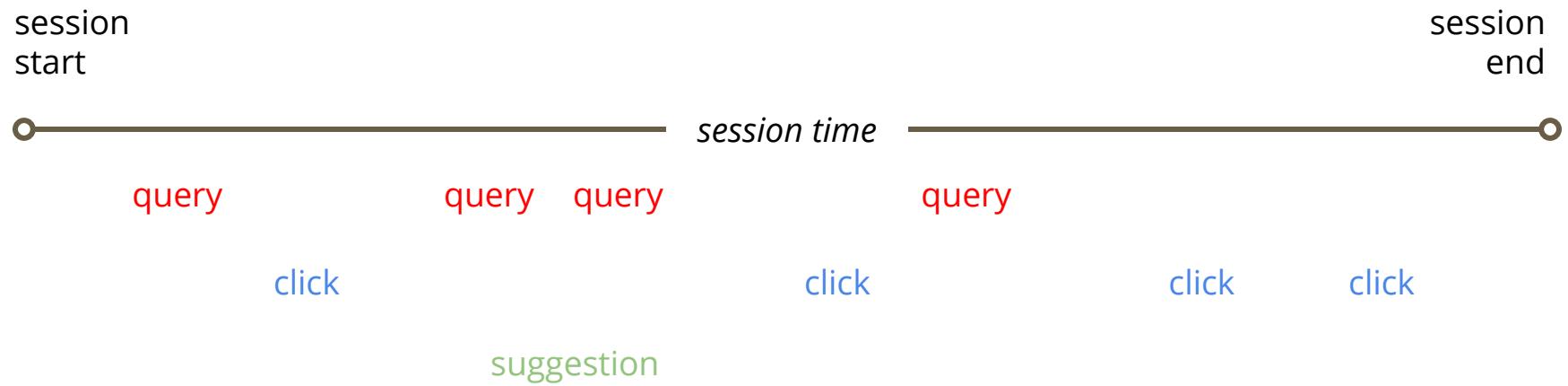
| | | |
|------------|-----------------------|---|
| 5:54:51 PM | Query | career development advice |
| 5:55:03 PM | └ Click | http://www.soperarticles.com/business-articles/career-devel... |
| 5:55:48 PM | Query | employment issues articles |
| 5:55:52 PM | └ Click | http://jobseekeradvice.com/category/employment-issues/... |
| 6:01:02 PM | Query | professional career advice |
| 6:01:05 PM | └ Click | http://ezinearticles.com/?Career-Advice-and-Professional-Ment... |
| 6:03:09 PM | └ Click | http://askville.amazon.com/buy-version-Tax-Software-year/Answer... |
| 6:03:35 PM | Query | what is a resume |
| 6:04:21 PM | └ Click | http://en.wikipedia.org/wiki/R%C3%A9sum%C3%A9 |
| 6:07:15 PM | END OF SESSION | |

exploring

| | | |
|------------|-----------------------|---|
| 1:20:15 PM | Query | can you use h & r block software for more than one year |
| 1:20:55 PM | Query | how do I file 2012 taxes on hr block |
| 1:20:58 PM | └ Click | http://www.hrblock.com |
| 1:33:17 PM | Query | can you only use h & r block one year |
| 1:33:29 PM | └ Click | http://www.consumeraffairs.com/finance/hr_block_free.html |
| 1:34:21 PM | └ Click | http://financialsoft.about.com/od/taxcut/gr/HR-Block-At-Home-... |
| 1:36:23 PM | Query | do I have to buy new tax software every year |
| 1:36:38 PM | └ Click | http://financialsoft.about.com/od/simpletips/f/upgrade_yearly.htm... |
| 1:55:10 PM | └ Click | http://askville.amazon.com/buy-version-Tax-Software-year/Answer... |
| 1:55:32 PM | END OF SESSION | |

struggling

Predicting Task Success



Rosie Jones and Kristina Lisa Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In Proceedings of the 17th ACM conference on Information and knowledge management (CIKM '08), 2008.

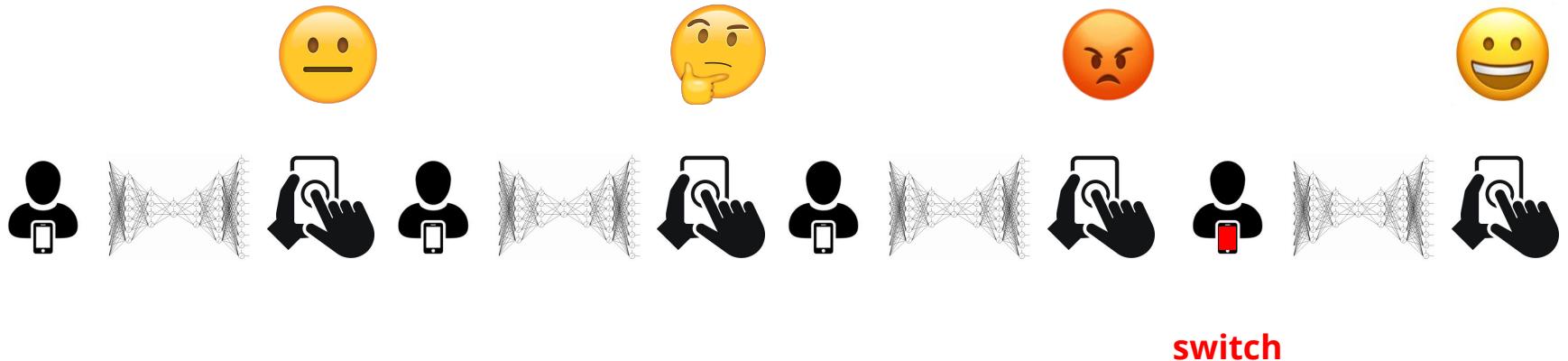
Task Success Signals

| Action | Description |
|-------------|--|
| Click_algoX | Click on the X-th algorithmic result |
| Click_Ans | Click on any answer (non-image) result |
| Click_IMG | Click on any image result |
| MouseRead | horizontal line across a result snippet of length > 50px and duration > 100 ms that goes from left to right which starts and ends inside an algo-result, or advertisement or an answer result |
| Scroll | page scroll recorded on the search engine result page |
| Move | any cursor movement of length > 10px and duration greater than > 50 ms |
| pause | smallPause: no cursor movement on the SERP for time < 5 seconds mediumPause: no cursor movement on the SERP for 5s < time < 20s longPause: no cursor movement on the SERP for 20s < time < 40s veryLongPause: no cursor movement on the SERP for time > 40s |
| Resize | change in the size of the window/screen encompassing the result page |
| IssueQuery | user movement to the Search Box on the SERP and typing of text in the query box |
| dwellTime | smallDwellTime: dwell time on a clicked result URL with time spent < 10s mediumDwellTime: dwell time on a clicked result URL with 10s < time < 40s longDwellTime: dwell time on a clicked result URL with time spent > 40s |
| QuickBack | click on a SERP URL followed by returning back to the SERP within 5s |

Task Success Signals (speech)

| Speech Features | |
|---------------------|---|
| PhoneticSim | Levenstein distance between Metaphone representation of the two queries |
| SRConf | Speech recognition confidence |
| LMScore | Language model score |
| SRRecoTime | Speech recognition time |
| Q2NBestExactMatch | Q2 is in N best candidate recognitions of Q1 (exact match) |
| Q2NBestPartialMatch | Q2 is in N best candidate recognitions of Q1 (partial match) |
| DeltaWordRate | Diff. in speech rate (words/sec) |
| DeltaLoudness | Diff. in voice loudness |
| DeltaPitch | Diff. in voice pitch |

Task Success Signals (switching)



Task Success Signals (switching)

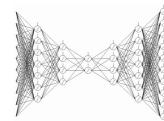
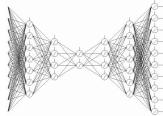


off-service instrumentation for switch detection

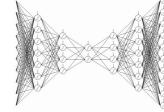
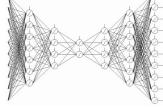
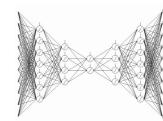
- adjacent
 - toolbar
 - messenger
- platform
 - browser
 - OS
 - conversational agent
- partner

Common Behaviors

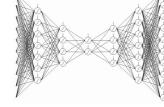
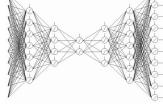
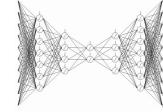
| behavior | question | example |
|--------------|--|--|
| attention | did the user see the system decision? | <i>page load</i> |
| interaction | did the user interact with the recommendation? | <i>track stream</i> |
| task success | was the user satisfied by the recommendation? | <i>completed playlist</i> |
| retention | is this a useful tool for the user? | <i>return to recommendation system</i> |



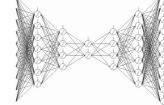
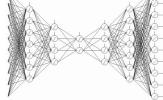
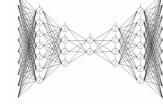
...



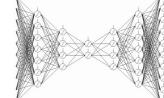
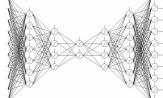
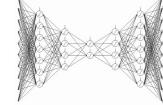
...



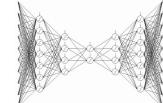
...



...



...



Search Engine Switching Signals

| Feature | Description |
|----------------------|--|
| fractionEngine | Fraction of queries issued to search engine |
| queryCountEngine | Number of queries issued to search engine |
| avgEngineQueryLength | Average length (in words) of queries to search engine |
| fractionEngineSAT | Fraction of search engine queries that are satisfied |
| fractionNavEngine | Fraction search engine queries defined as navigational |
| fractionNavEngineSAT | Fraction of queries in fractionNavEngine satisfied |

The Importance of Logging

- models of satisfaction based on observed user behavior are only as good as the logging mechanism
 - unlogged behavior cannot be modeled
 - logging mechanisms are the eyes and ears of an online, adaptive recommender system

Data Corruption

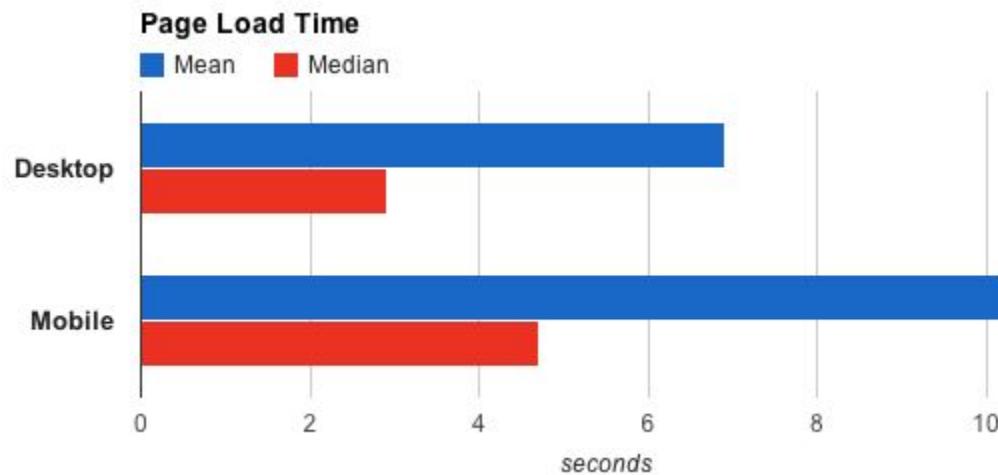
- logging noise
 - technical issues (e.g. optimizing away beacons)
 - poor randomization
 - interacting instrumentation (e.g. cursor-tracking increases page load time)
- audit behavioral instrumentation with data quality metrics
 - A/A test comparing baseline metrics before and after instrumentation
 - test with simulated user behavior

Z. Zhao, M. Chen, D. Matheson and M. Stone, "Online Experimentation Diagnosis and Troubleshooting Beyond AA Validation," DSAA 2016.

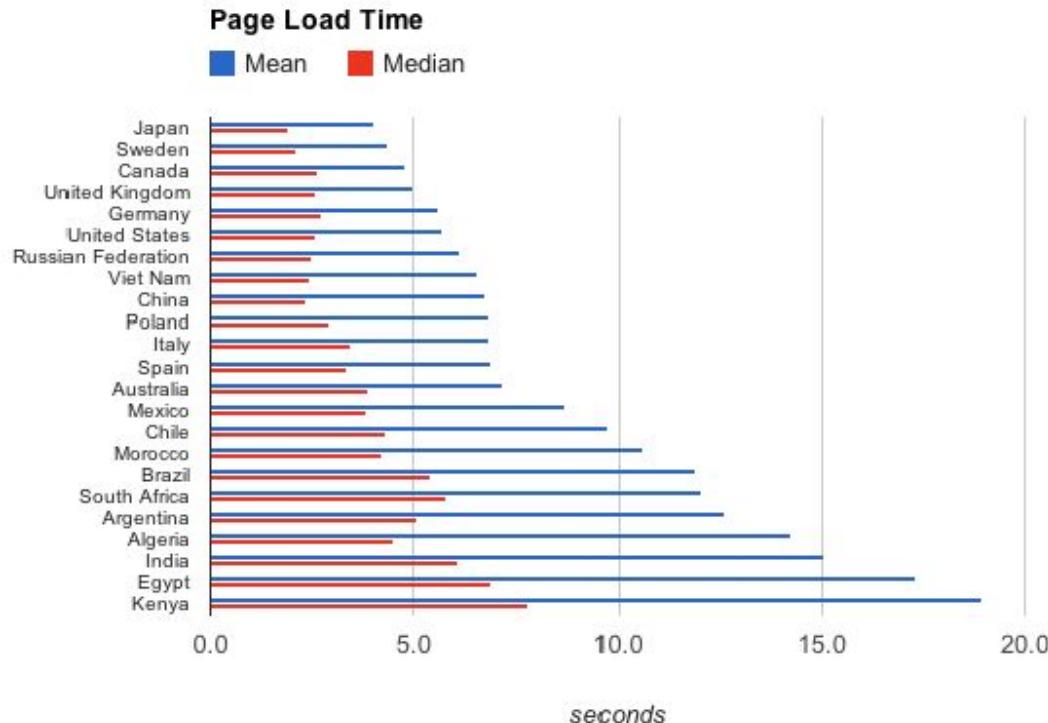
Pavel Dmitriev, Somit Gupta, Dong Woo Kim, and Garnet Vaz. Twelve Common Metric Interpretation Pitfalls in Online Controlled Experiments. KDD 2017.

Thomas Crook, Brian Frasca, Ron Kohavi, and Roger Longbotham. Seven pitfalls to avoid when running controlled experiments on the web. KDD, 2009.

Data Bias: Dwell Time



Data Bias: Dwell Time



Data Bias: Dwell Time

| Character Size (deg) | Young (n = 16) Rate | Old (n = 14) | | Old-Normal (n = 9) | |
|-------------------------|---------------------------|--------------|------------------|--------------------|------------------|
| | | Rate | (%) ^a | Rate | (%) ^a |
| .15 | 258.59 | 170.87 | 66.1 | 190.32 | 73.6 |
| .30 | 314.47 | 284.68 | 90.5 | 304.60 | 96.9 |
| .50 | 349.18 | 294.84 | 84.4 | 310.95 | 89.1 |
| 1.00 | 343.21 | 291.90 | 85.0 | 306.38 | 89.3 |
| 4.00 | 240.34 | 186.69 | 77.7 | 191.31 | 79.6 |
| 12.00 | 163.51 | 121.28 | 74.2 | 124.13 | 75.9 |

^aPercentage of the young group's rate.

Data Bias: Visual Search

Table 3. Average Number of Center Task Misses Before and After Training

| Center Task Demand | Number of Distractors | Average Number of Center Task Misses | | | | | |
|--------------------------|--------------------------|--------------------------------------|--------------|-------------|--------------|-------------|--------------|
| | | Young | | Middle Aged | | Old | |
| | | Pretraining | Posttraining | Pretraining | Posttraining | Pretraining | Posttraining |
| High | 47 | 11.375 | 5.000 | 18.625 | 10.250 | 18.375 | 14.500 |
| | 23 | 12.250 | 3.625 | 15.250 | 12.625 | 19.125 | 15.000 |
| | 0 | 11.250 | 4.875 | 14.000 | 10.750 | 22.000 | 13.000 |
| Medium | 47 | 3.625 | 0.125 | 6.000 | 1.625 | 8.625 | 5.250 |
| | 23 | 2.250 | 0.625 | 8.125 | 1.000 | 8.375 | 4.625 |
| | 0 | 3.000 | 1.250 | 5.250 | 0.375 | 7.125 | 4.250 |
| Low | 47 | 1.250 | 0.000 | 2.000 | 0.125 | 2.750 | 1.000 |
| | 23 | 0.375 | 0.000 | 0.625 | 0.375 | 3.875 | 0.875 |
| | 0 | 0.625 | 0.625 | 0.750 | 0.250 | 2.375 | 0.500 |

Ball, Beard, Roenker, Miller, Griggs. "Age and visual search: expanding the useful field of view". Journal of the Optical Society of America A, 1988.

New Domains

- new domains require new signals
 - music
 - movies
- bootstrapping new signals
 - literature review
 - pilot experiments



**quantitative behavioral signals should be developed
and vetted in collaboration with qualitative research**

Best Practices for Instrumentation

respect for persons: protect autonomy of users; informed consent

beneficence: minimize risk for users

justice: fair distribution of risk

The belmont report: ethical principles and guidelines for the protection of human subjects of research. Dept. of Health, Education and Welfare, Office of the Secretary, Washington D.C., 1988.

Luke Stark. Algorithmic psychometrics and the scalable subject. Social Studies of Science, 48(2):204-231, 2018.

future directions

new modalities and platforms (e.g. audio-only, virtual, IoT)

transparency of behavioral logging

privacy-sensitive logging

Break

Coming up at 11:00am:

- Qualitative data analysis
 - Quantitative data analysis
 - Best practices
-