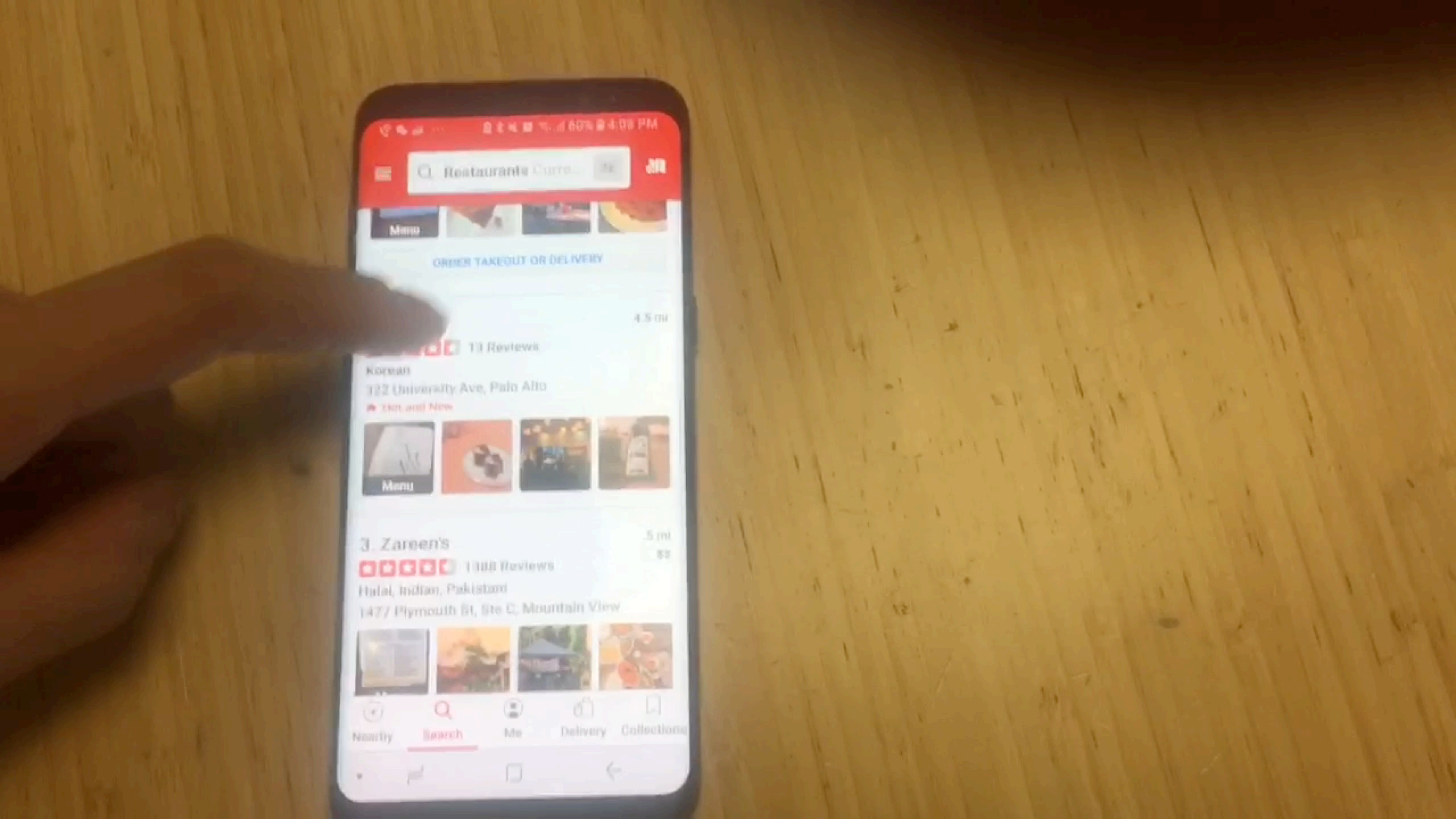
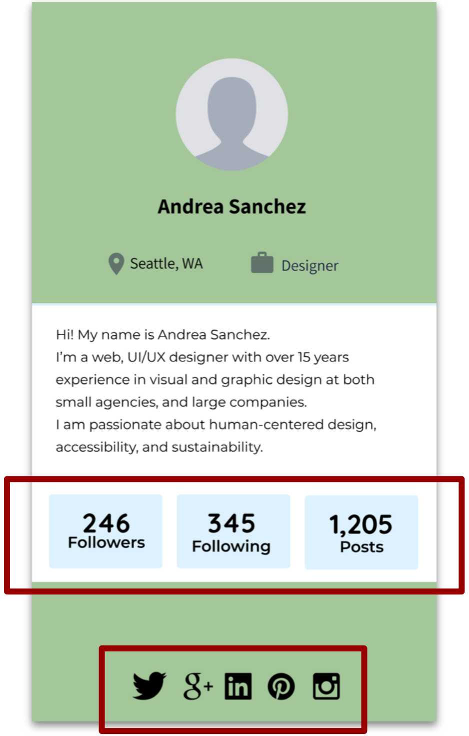
Modeling Mobile Interface Tappability Using Crowdsourcing and Deep Learning

Amanda Swearngin, University of Washington

Yang Li, Google


Restaurants
Menu
ORDER TAKEOUT OR DELIVERY
Reviews
Korean
University Ave, Palo Alto
Hot and New
Menu
3. Zareen's
Reviews
Halal, Indian, Pakistani
Plymouth St, Ste C, Mountain View
Nearby
Search
Me
Delivery
Collections

Signifiers



Color (e.g., A Link)

Shape of a button

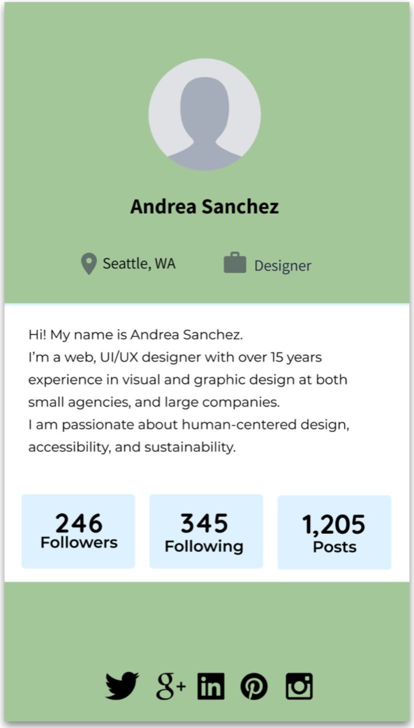
Location on screen

A combination

…

Don Norman, "The Way I See It: Signifiers, Not Affordances", Interactions, 2008

What if a designer uses the wrong signifier?



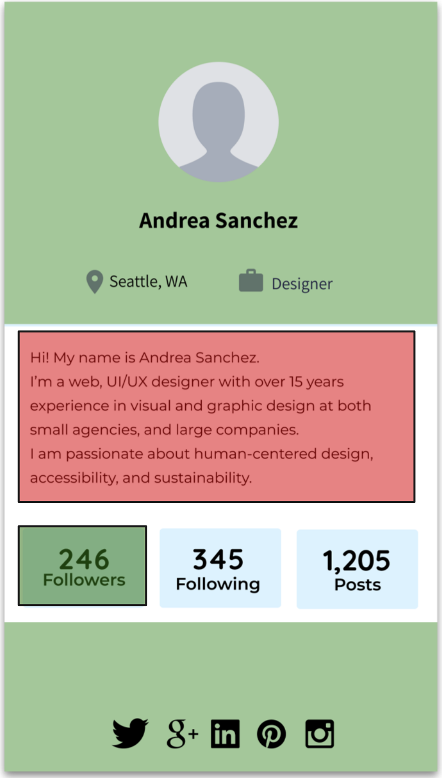
No Signifier

- Lack of discoverability

False Signifier

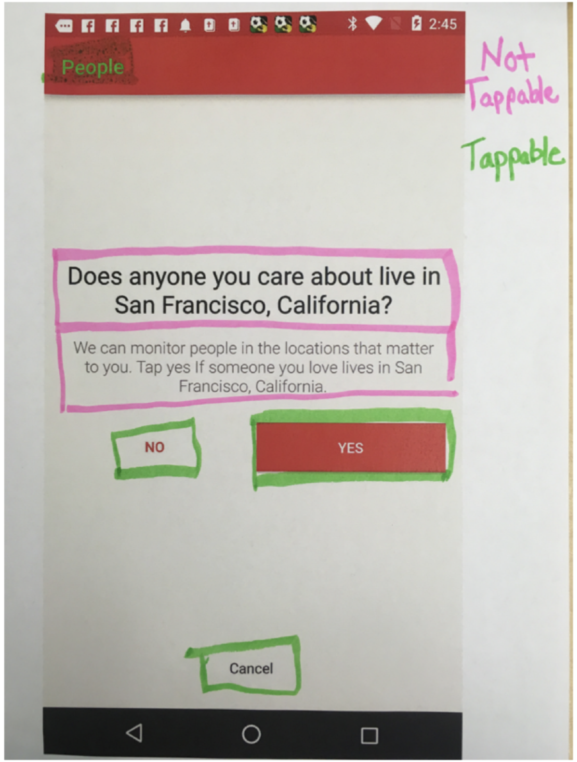
- Tap with no response -> frustration

A Tappability Study



Tappable

Not Tappable



Challenges

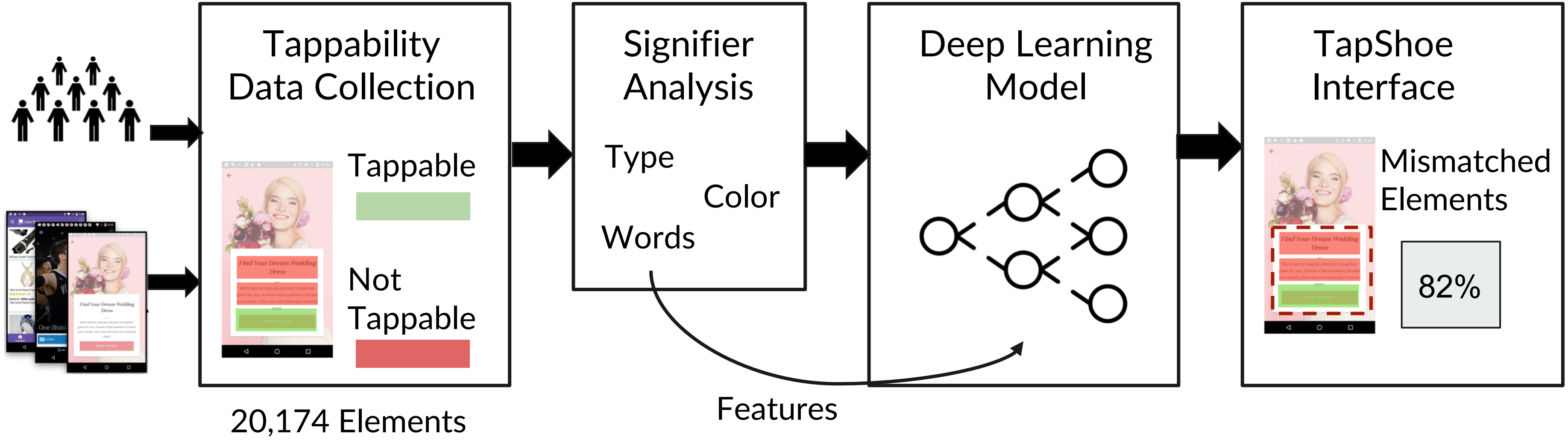
For Designers

- Expensive
- Time consuming

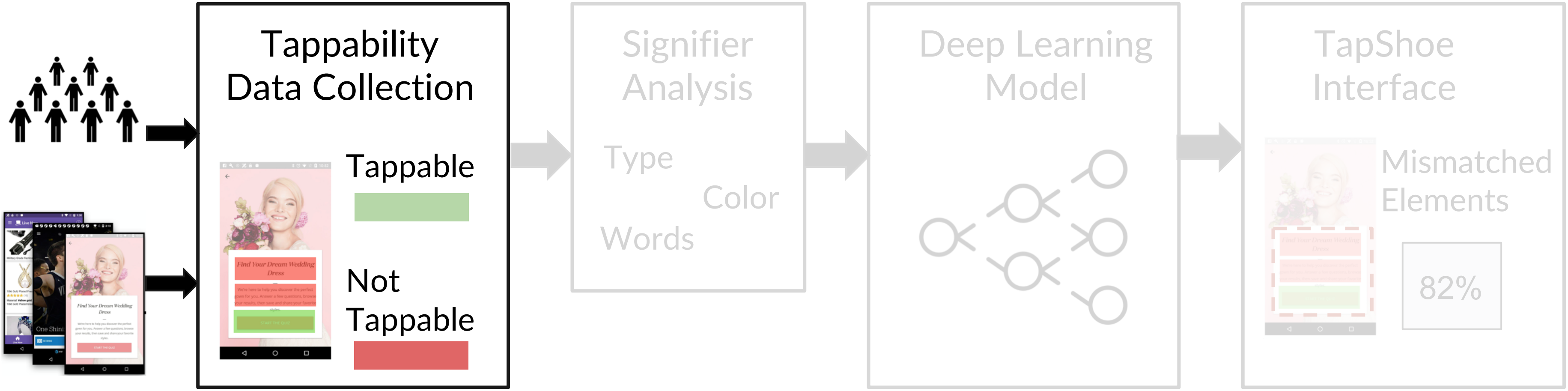
For design & research community

- No consistent understanding of signifiers at a large scale.
- Diverse tappability data needed to build automated approaches.

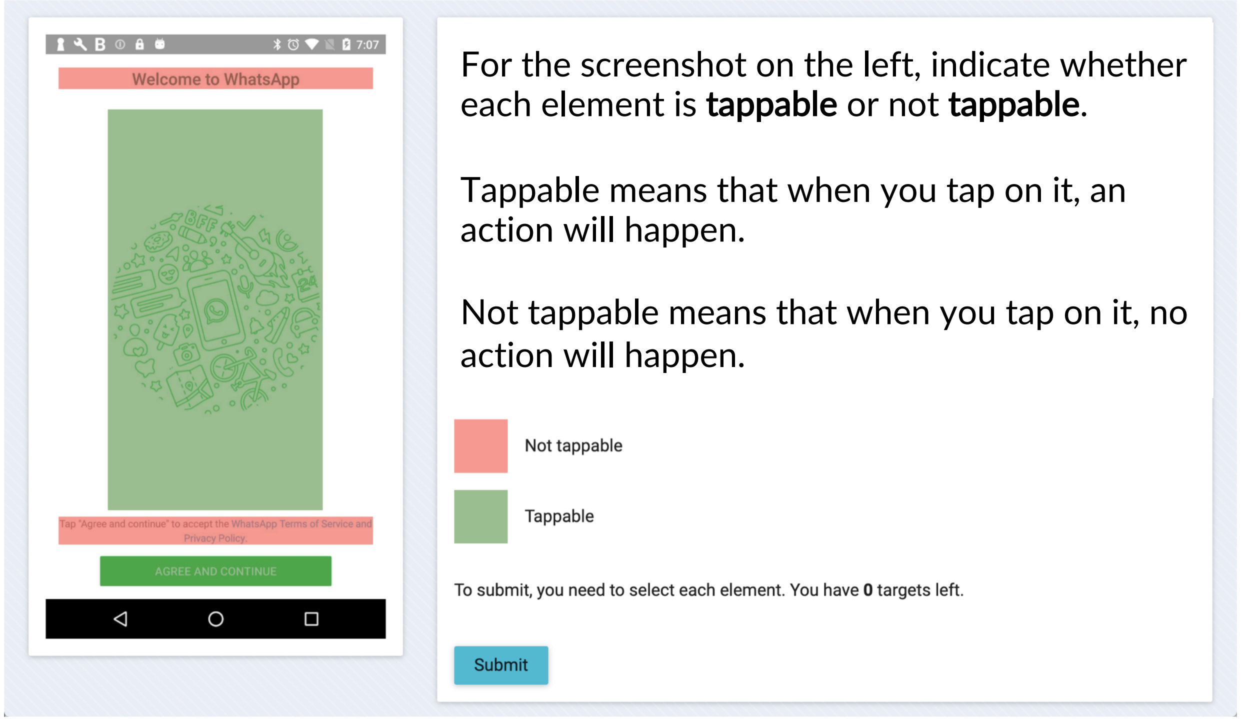
Our Approach



Talk Overview



Crowdsourcing Tappability Data



Tappability Data

3,470 screens

743 workers

20,174 elements

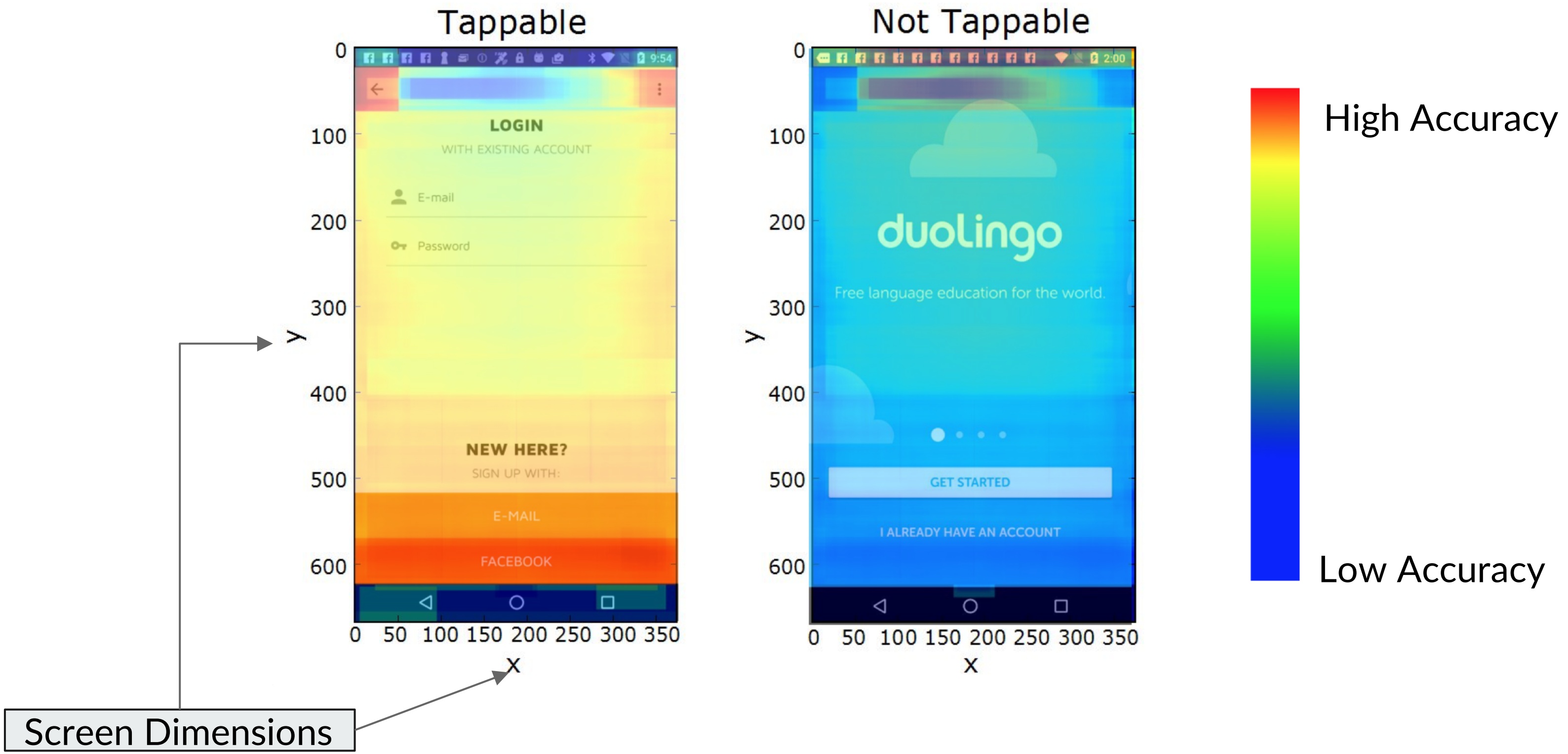
Deka, et, al., "Rico: A mobile app dataset for building data-driven design applications", UIST 2016

Accuracy of the Worker Labels

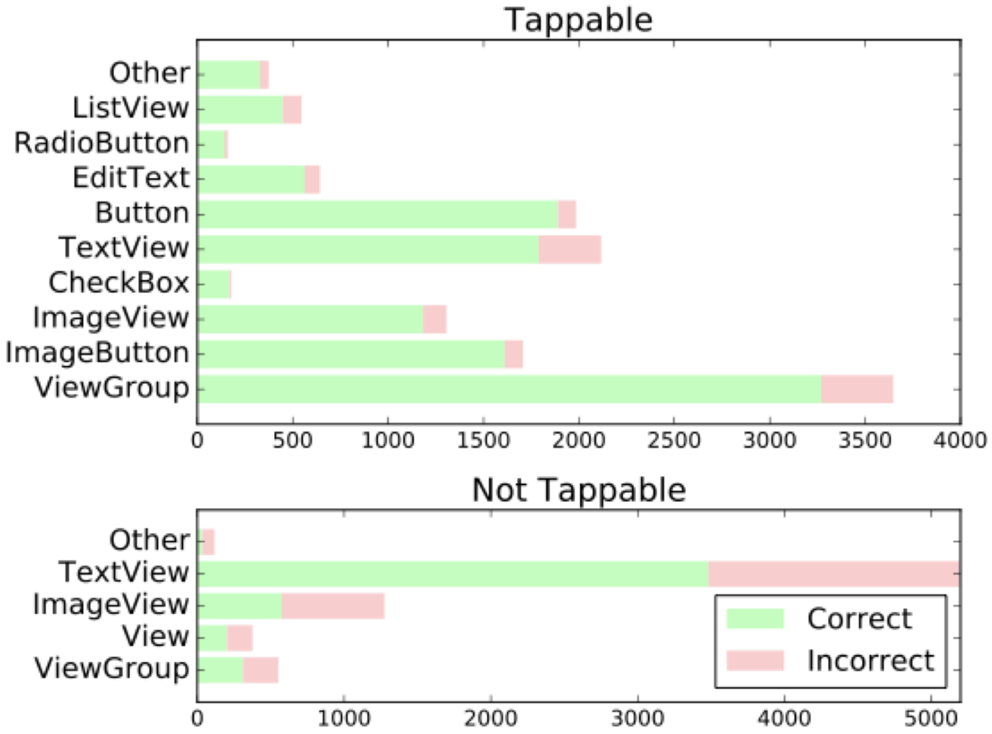
	# Labels Collected	Precision	Recall
Tappable	14,301	89.99%	79.67%
Not Tappable	5,873	61.31%	78.43%

Talk Overview

How does location indicate tappability?



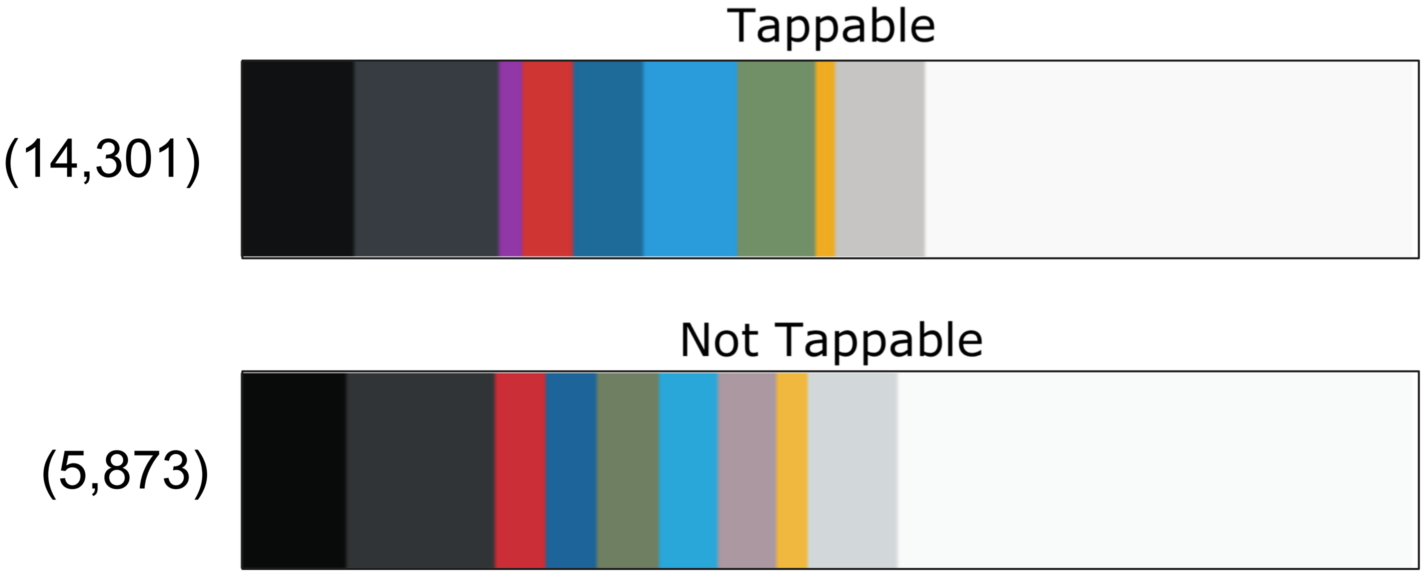
How does element type indicate tappability?



Tappable elements more correct, especially for common tappable elements (e.g., buttons, checkboxes)

Not tappable elements most common types have more flexibility in design -> more ambiguity.

What colors are more common in tappable elements?



Beyond Blue Links: Making Clickable Elements Recognizable.,
https://www.nngroup.com/articles/clickable-elements/, 2015

Do tappable elements have fewer words, and more actionable keywords?

Not tappable elements had **1.84 more words per element, on average.**

Top 5 **Tappable** Keywords ← TF-IDF Analysis

1. **Submit**
2. **Close**
3. Brown
4. Grace
5. Beauty

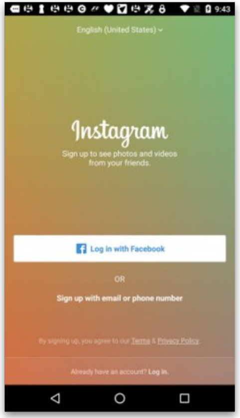
Jenifer Tidwell, Designing Interfaces: Patterns for Effective Interaction Design, 2015.

Talk Overview

Tappability Model

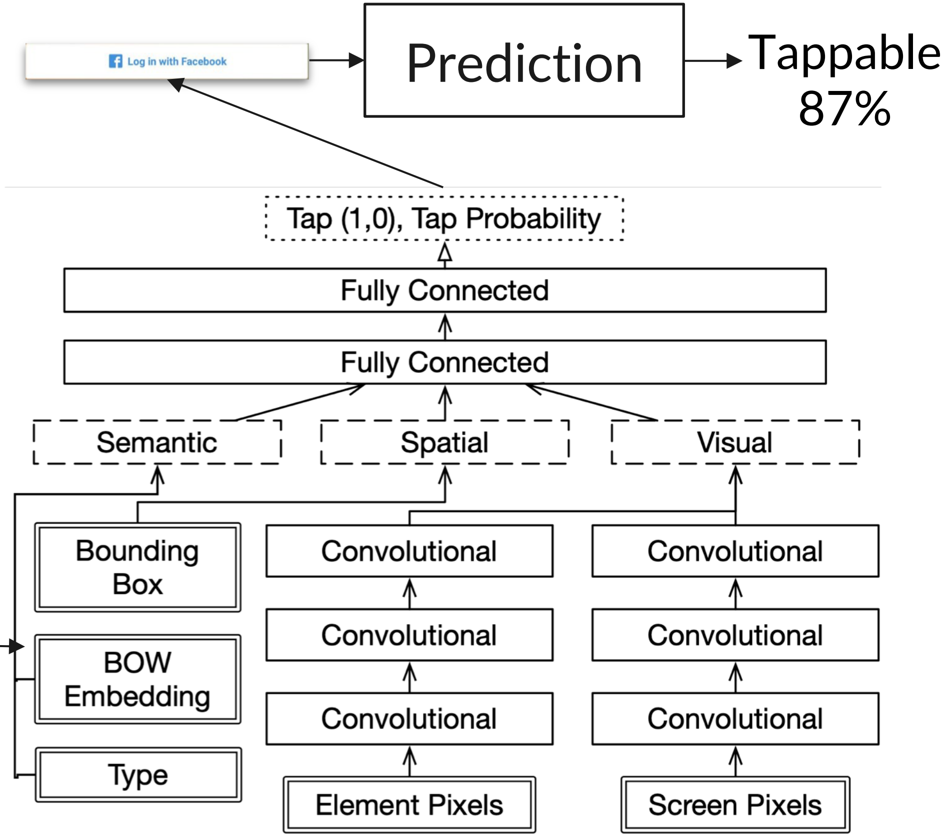
Features

Bounding Box
{ x: 22, y: 30, width: 120, height: 40}



Type: Button

“Log in with Facebook”,
Number of Words: 4

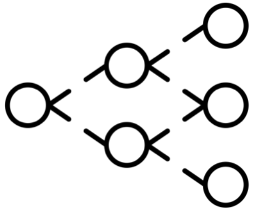


How well can we predict tappability?

	Tappable	Not Tappable
Original Dataset	P: **90.2% (SD: 0.3)** R: **87.0% (SD: 1.6)**	P: **70% (SD: 2.0)** R: **78% (SD: 3.0)**
Balanced Dataset	P: **82% (SD: 0.3)** R: **84% (SD: 1.6)**	P: **81% (SD: 2.0)** R: **86% (SD: 3.0)**

How can we improve the model's accuracy?

Add more features, improve model



Are human labels inconsistent?

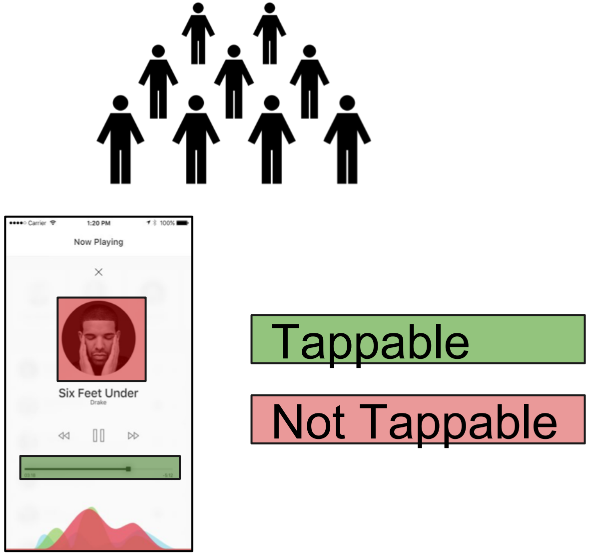
How consistent are the tappability labels?

290 workers

2,000 unique elements

334 screenshots

Each element labeled 5 times



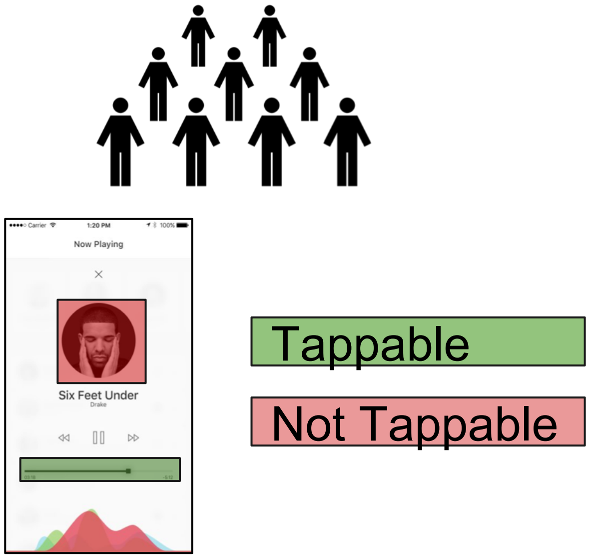
How consistent are the tappability labels?

Results

58% elements labeled the same among all 5 workers.

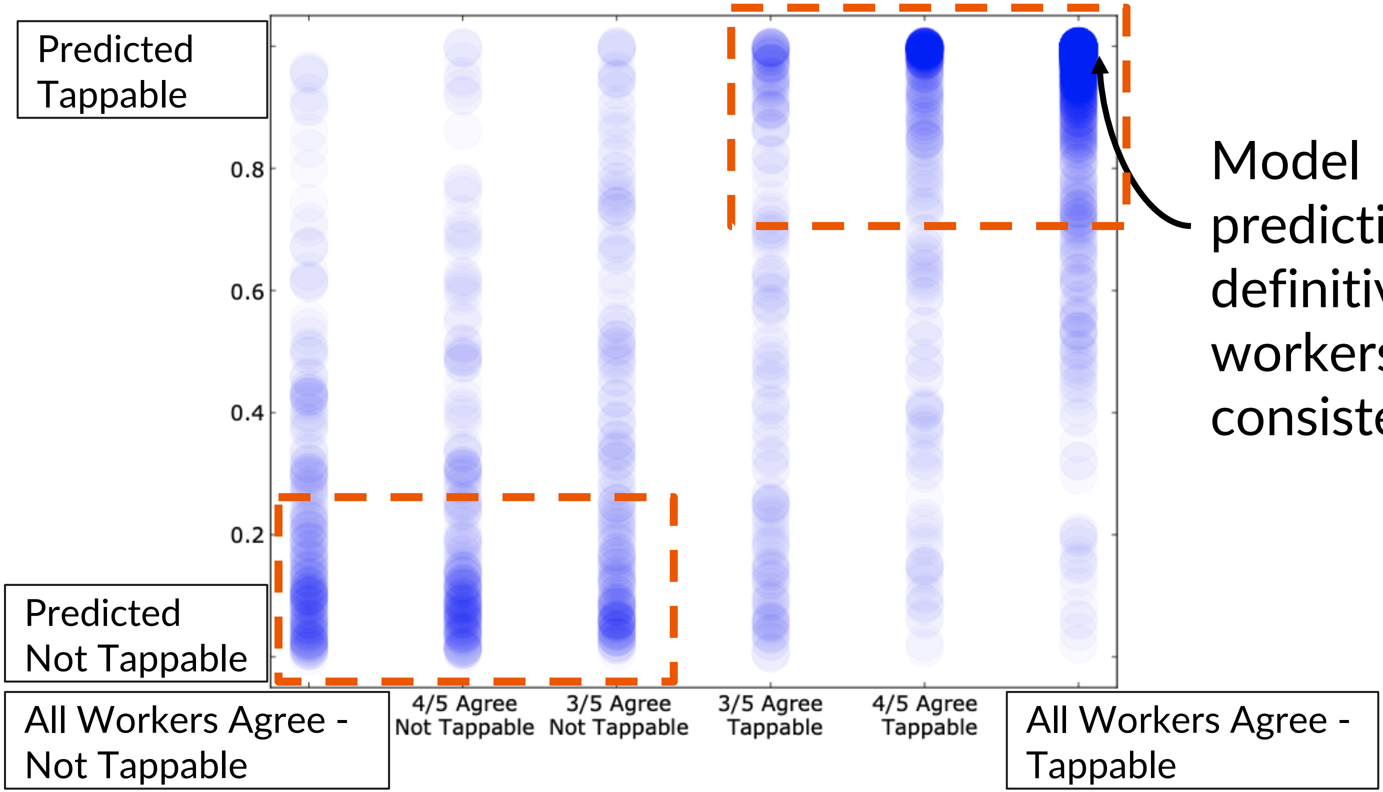
Agreement Score[1]: 0.834

Fleiss' Kappa[2]: 0.520 (Moderate)



1. Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. "Maximizing the Guessability of Symbolic Input", CHI 2005
2. Joseph L Fleiss, "Measuring Nominal Scale Agreement Among Many Raters", Psychological Bulletin, 1971

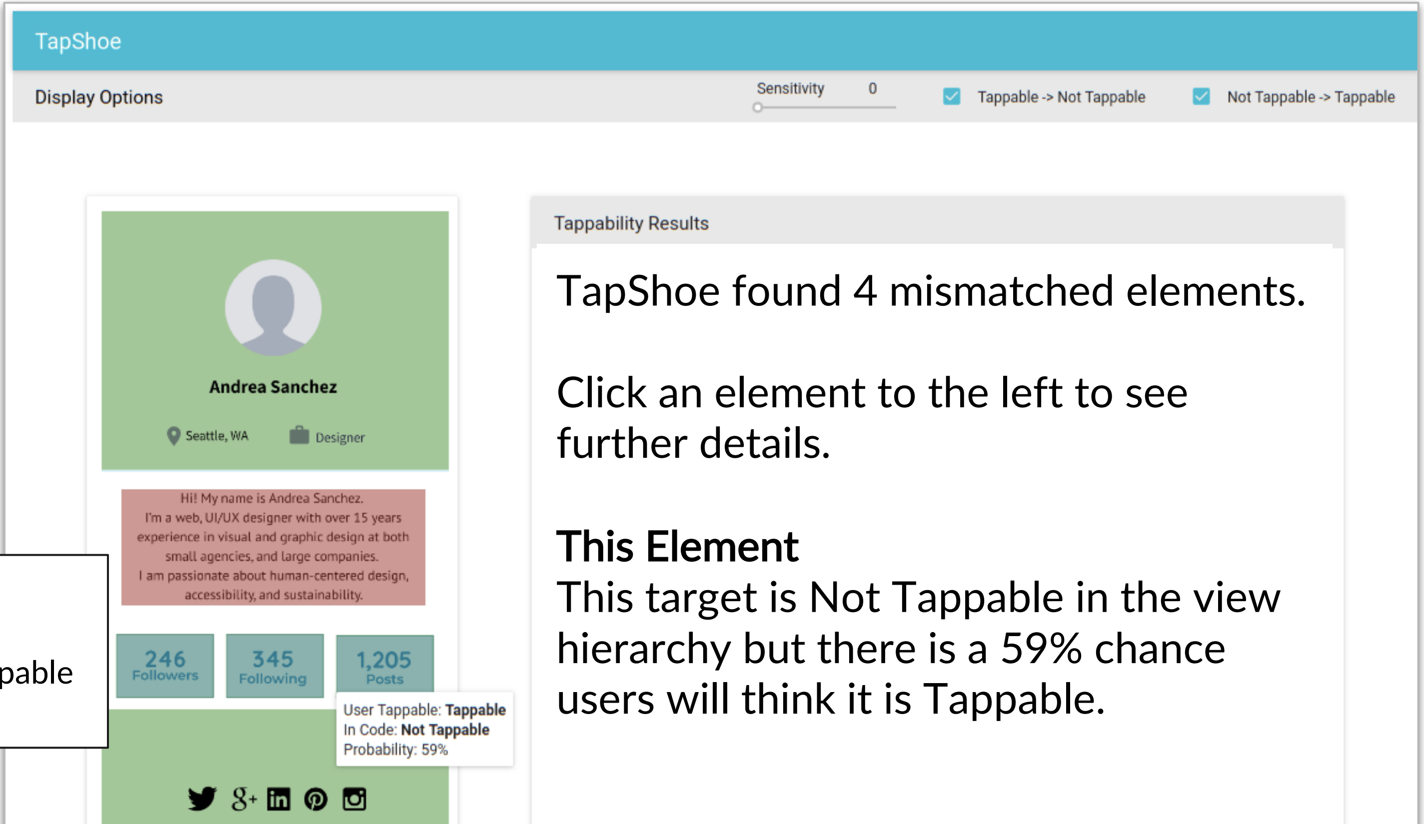
Do the model results reflect consistency?



Model predictions more definitive when workers are consistent.

Talk Overview

TapShoe Interface



User Tappable: Tappable
In Code: Not Tappable
Probability: 59%

Designer Interviews

Informal interviews with 7 professional designers

Demonstrated them TapShoe interface and model

Questions:

- How do you see the TapShoe interface fitting into your design process?
- How can you envision using the models predictions, beyond the TapShoe interface?

How can we help designers understand tappability?

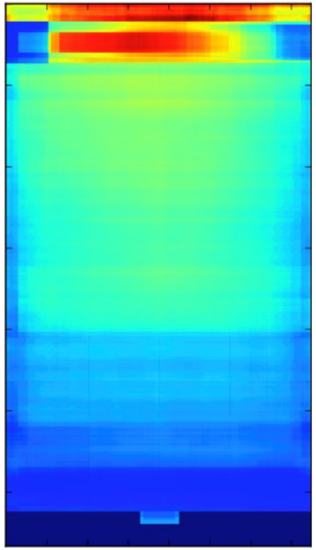
TapShoe interface - Provide recommendations for a fix

Spatial visualization of tappability (i.e., Heatmap)

Tool to explore small variations, and discover new signifers.

Train on existing datasets or platforms

Predictions on early stage mockups (i.e., Sketch documents)



Modeling Mobile Interface Tappability Using Crowdsourcing and Deep Learning

Key Takeaways:

- People have low accuracy in distinguishing tappable from not tappable elements.
- We can build models that use visual, spatial, and semantic features to predict human tappability perception.
- This can help designers understand and improve the usability of their interfaces.

Amanda Swearngin*
Paul G. Allen School
University of Washington
amaswea@cs.washington.edu

Yang Li
Google Research
yangli@acm.org

* This work was completed while the first author was an intern at Google.