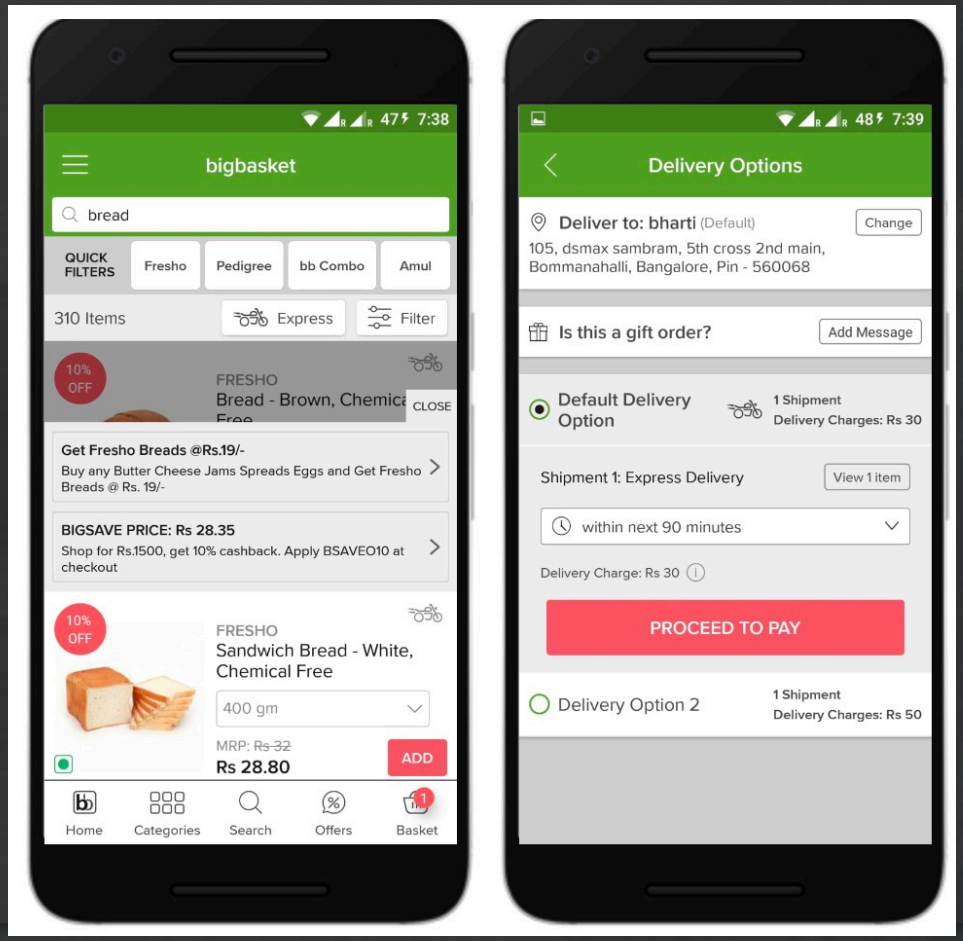
Heuristic Evaluation

刘哲明
Prof. James A. Landay
Computer Science Department
Stanford University

Winter 2022
February 14, 2022

Hall of Fame or Shame?



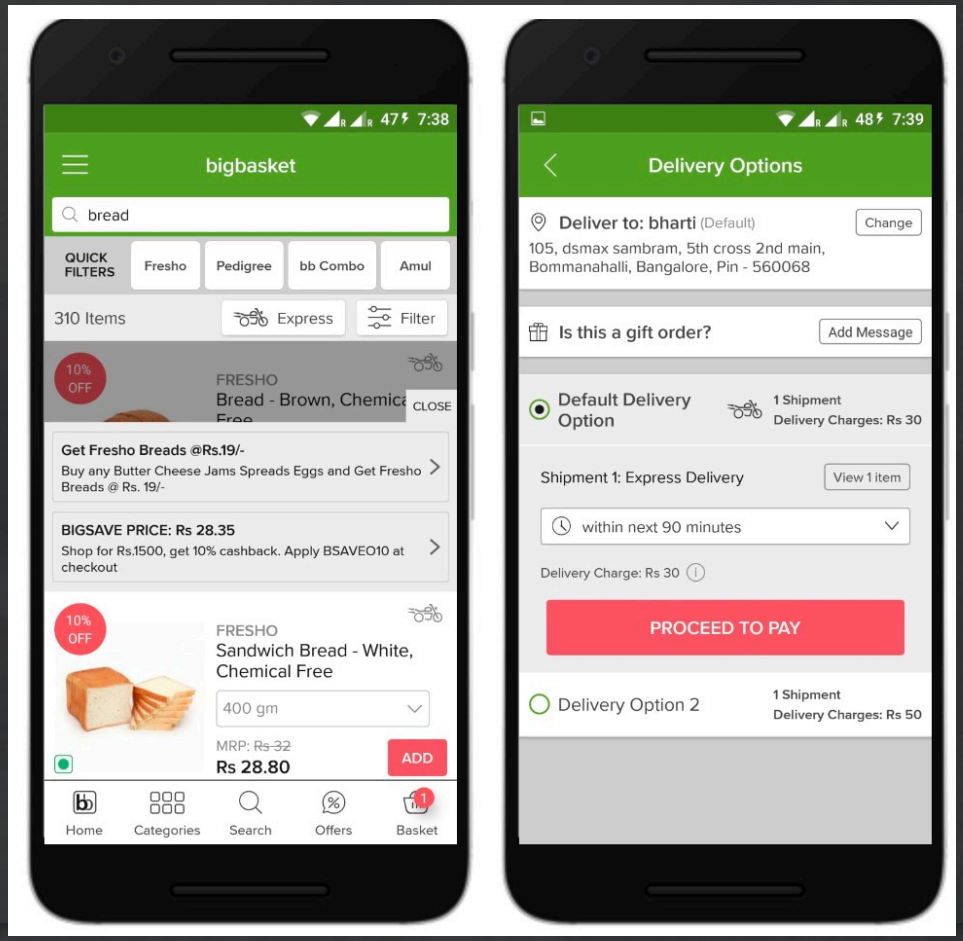
Big basket

From Bharti Bhagtani

https://uxdesign.cc/heuristic-evaluation-of-bigbasket-application-4a69f43be47d

"India's largest online supermarket"

Hall of Shame!



Big basket

From Bharti Bhagtani

https://uxdesign.cc/heuristic-evaluation-of-bigbasket-application-4a69f43be47d

"India's largest online supermarket"

Good

- ?

Bad

- not aesthetic & minimalist design
- popups with too much info
- cluttered

Heuristic Evaluation

刘哲明
Prof. James A. Landay
Computer Science Department
Stanford University

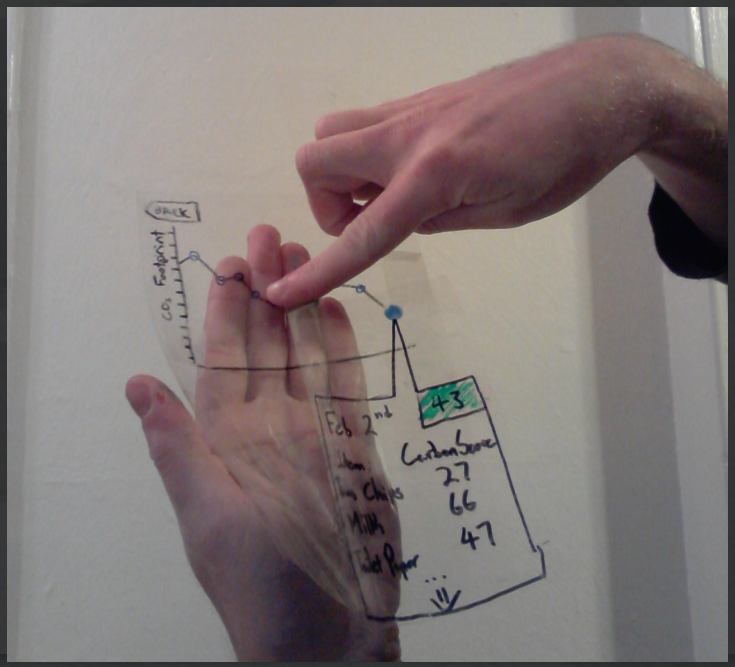
Winter 2022
February 14, 2022

Outline

- Heuristic Evaluation Overview
- The Heuristics
- Team Break
- Exercise

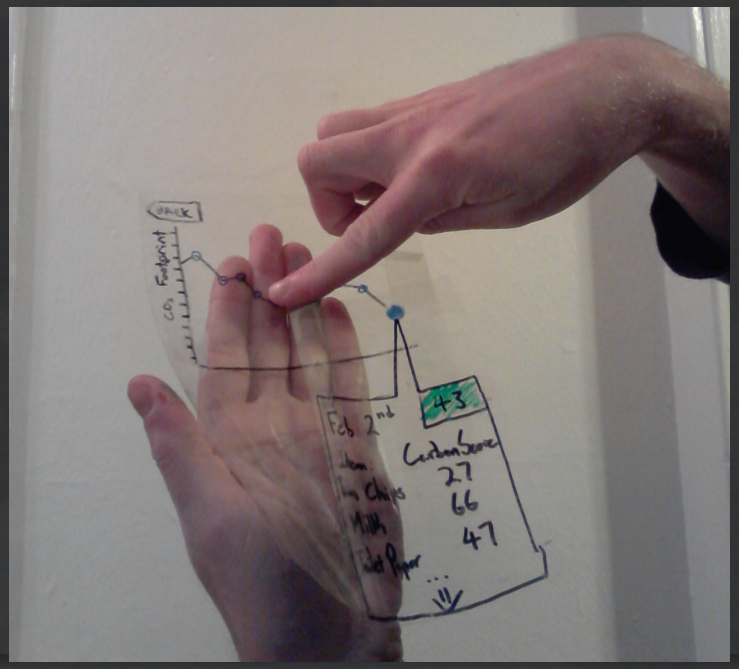
Evaluation

- About figuring out how to improve design
- Issues with lo-fi tests?



Evaluation

- About figuring out how to improve design
- Issues with lo-fi tests?



Not realistic
– visuals & performance

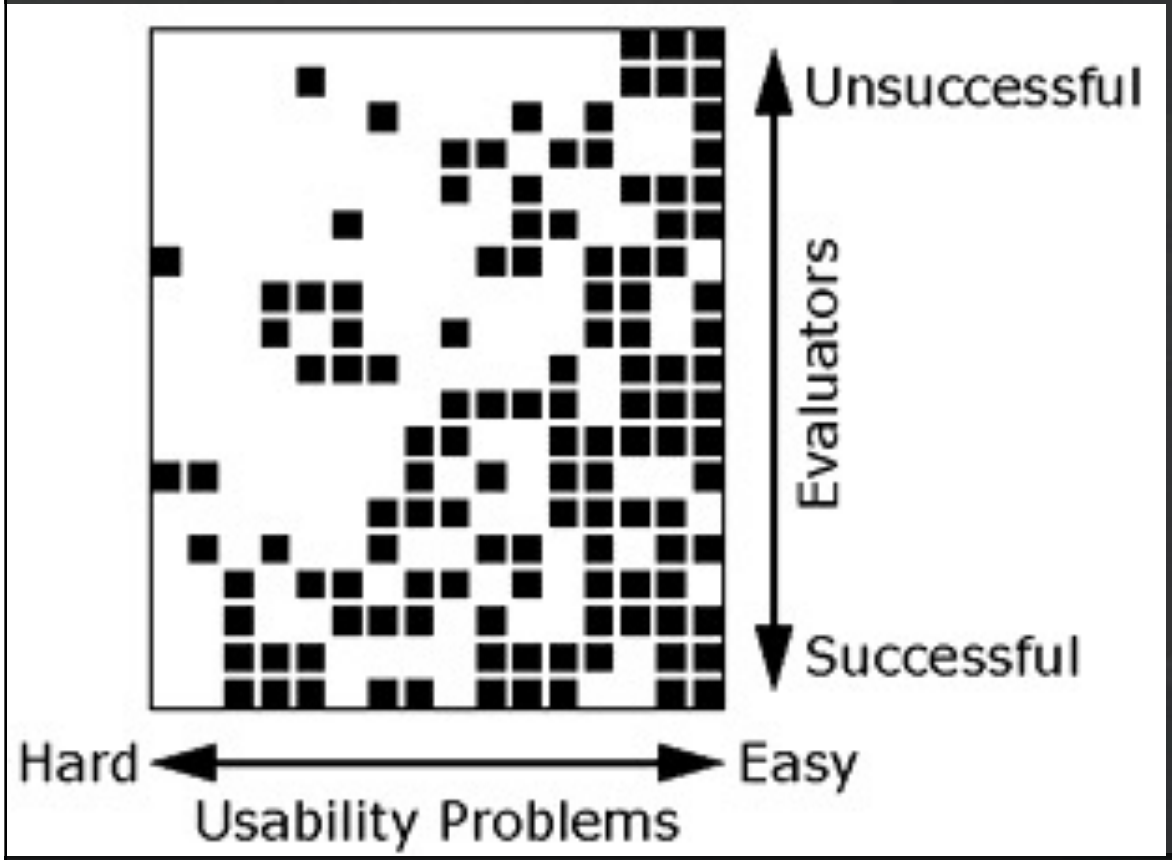
Not on actual interface
– can't test alone

Need participants
– can be hard to find repeatedly

Heuristic Evaluation

- Developed by Jakob Nielsen
- Helps find usability problems in a UI design
- Small set (3-5) of evaluators examine UI
  - independently check for compliance with usability principles (“heuristics”)
  - evaluators only communicate afterwards
    - findings are then aggregated
  - use violations to redesign/fix problems
- Can perform on working UI or on sketches

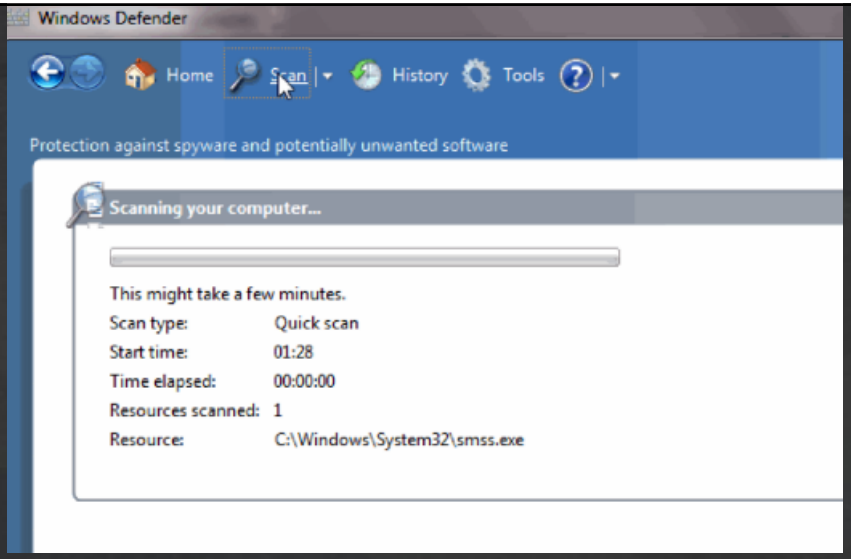
Why Multiple Evaluators?

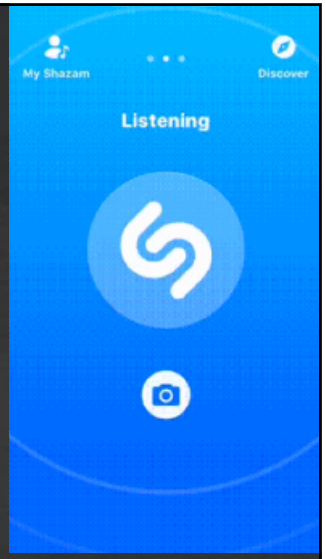


- Every evaluator doesn't find every problem
- Good evaluators find both easy & hard ones

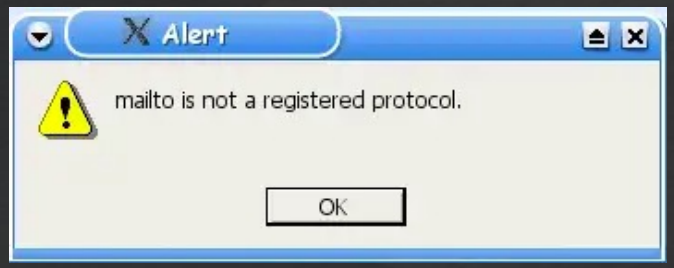
Heuristics

H1: Visibility of system status

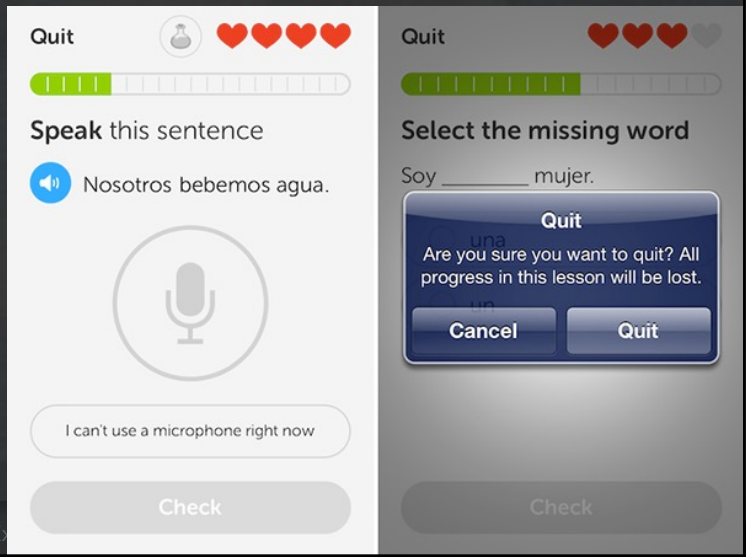




H2: Match between system & real world

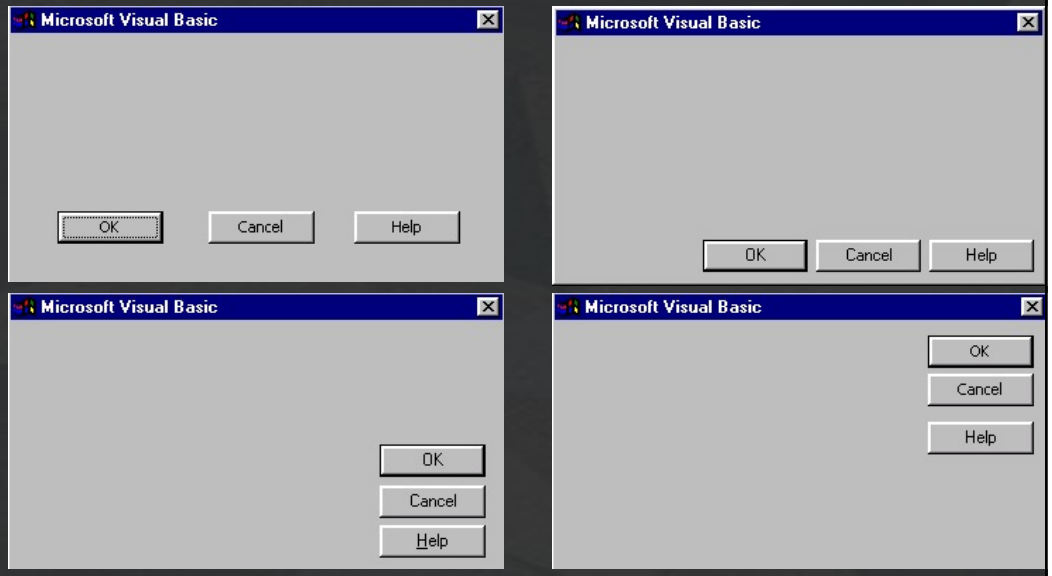


H3: User control & freedom



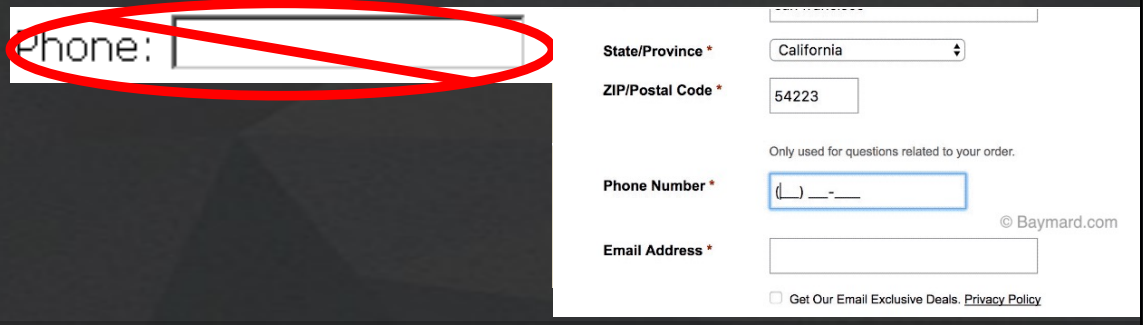
Heuristics (cont.)

H4: Consistency & standards



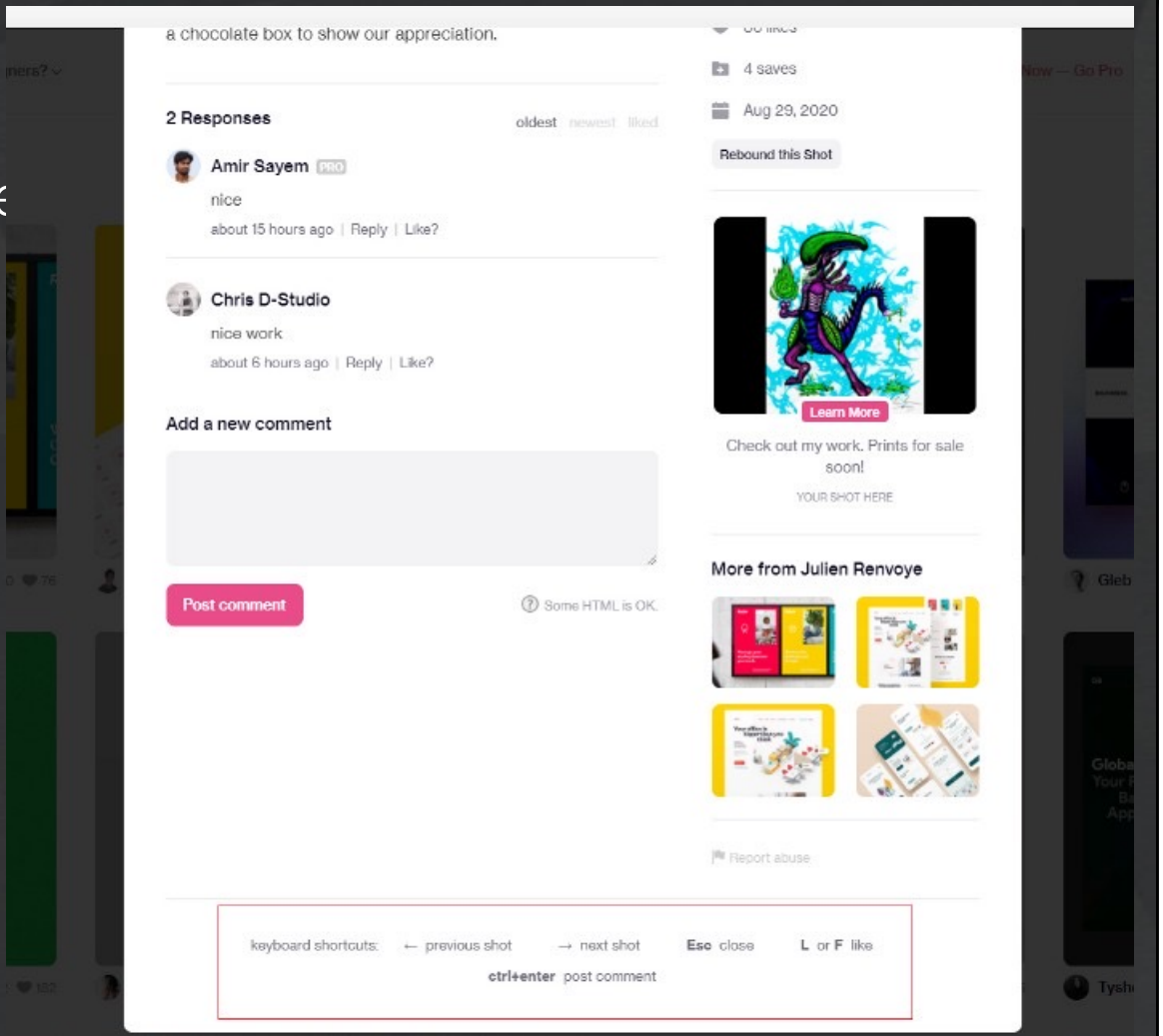
H5: Error prevention

H6: Recognition rather than recall



Heuristics (cont.)

H7: Flexibility and efficiency of use



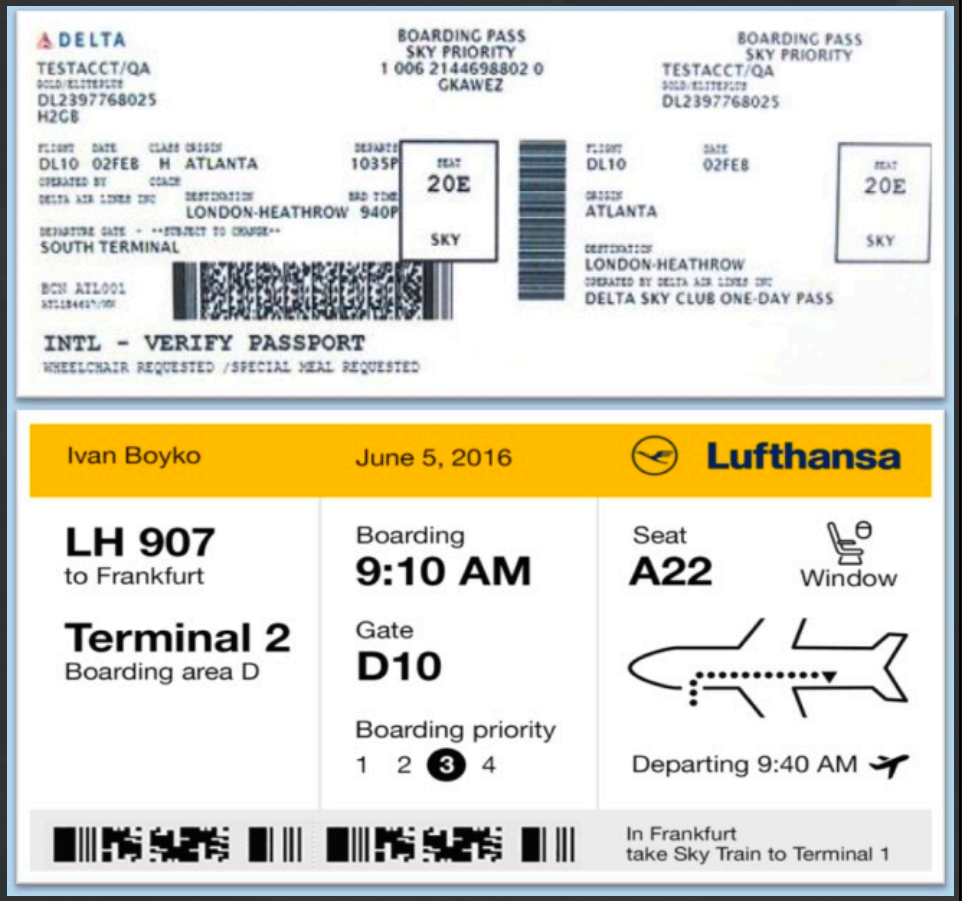
https://uxplanet.org/heuristics-7-flexibility-and-efficiency-of-use-simplified-by-the-examples-f2d76966e8a5

Heuristics (cont.)

H8: Aesthetic & minimalist design

bad

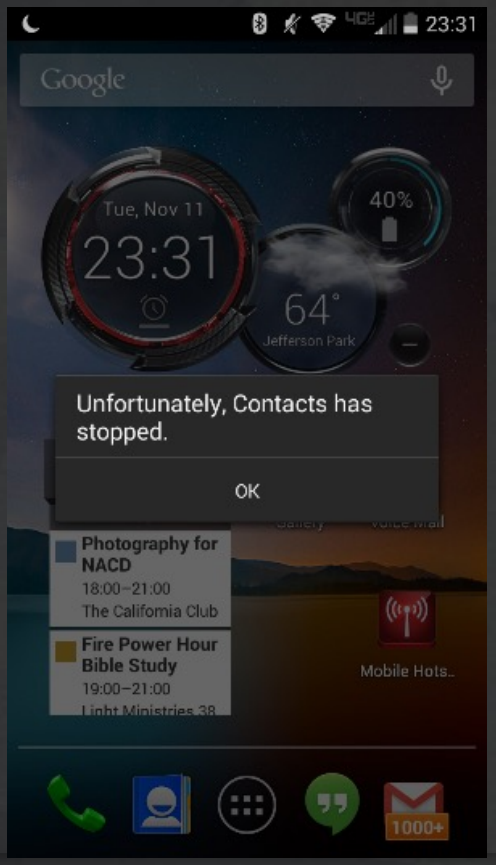
good



https://icons8.com/articles/redesigning-boarding-pass-again/

Heuristics (cont.)

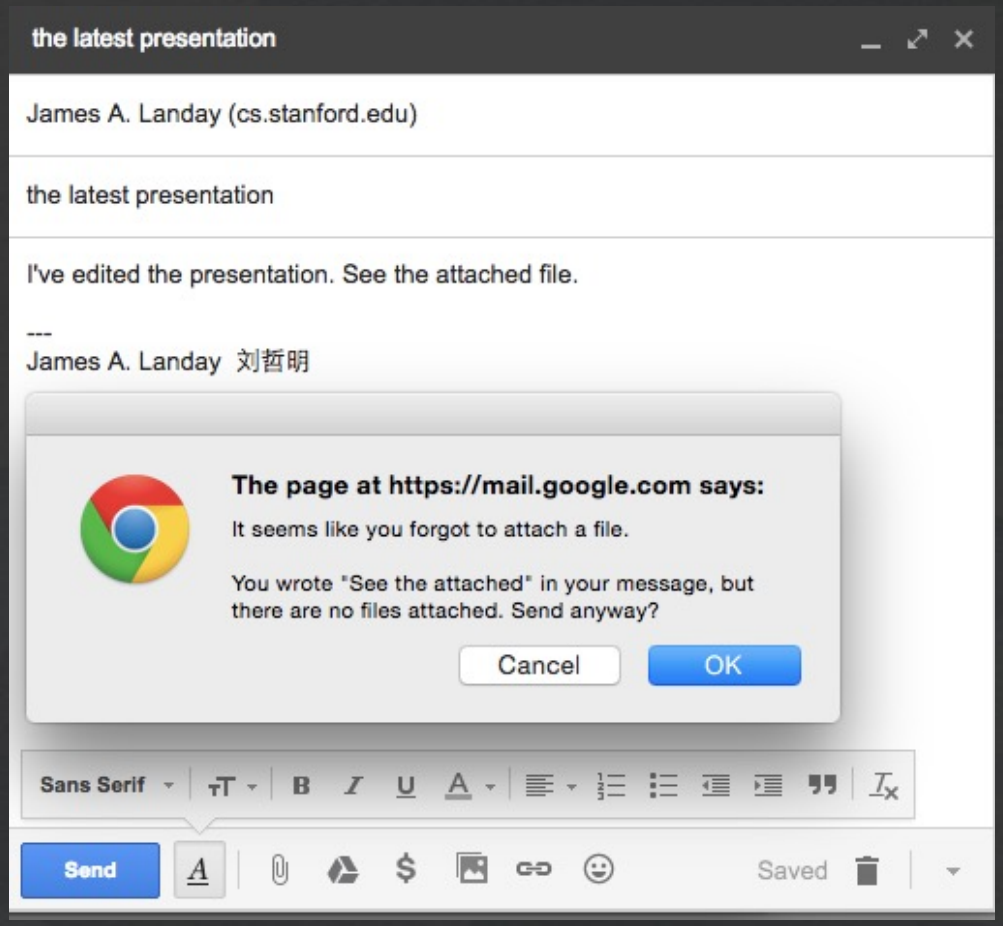
bad



H9: Help users recognize, diagnose, & recover from errors

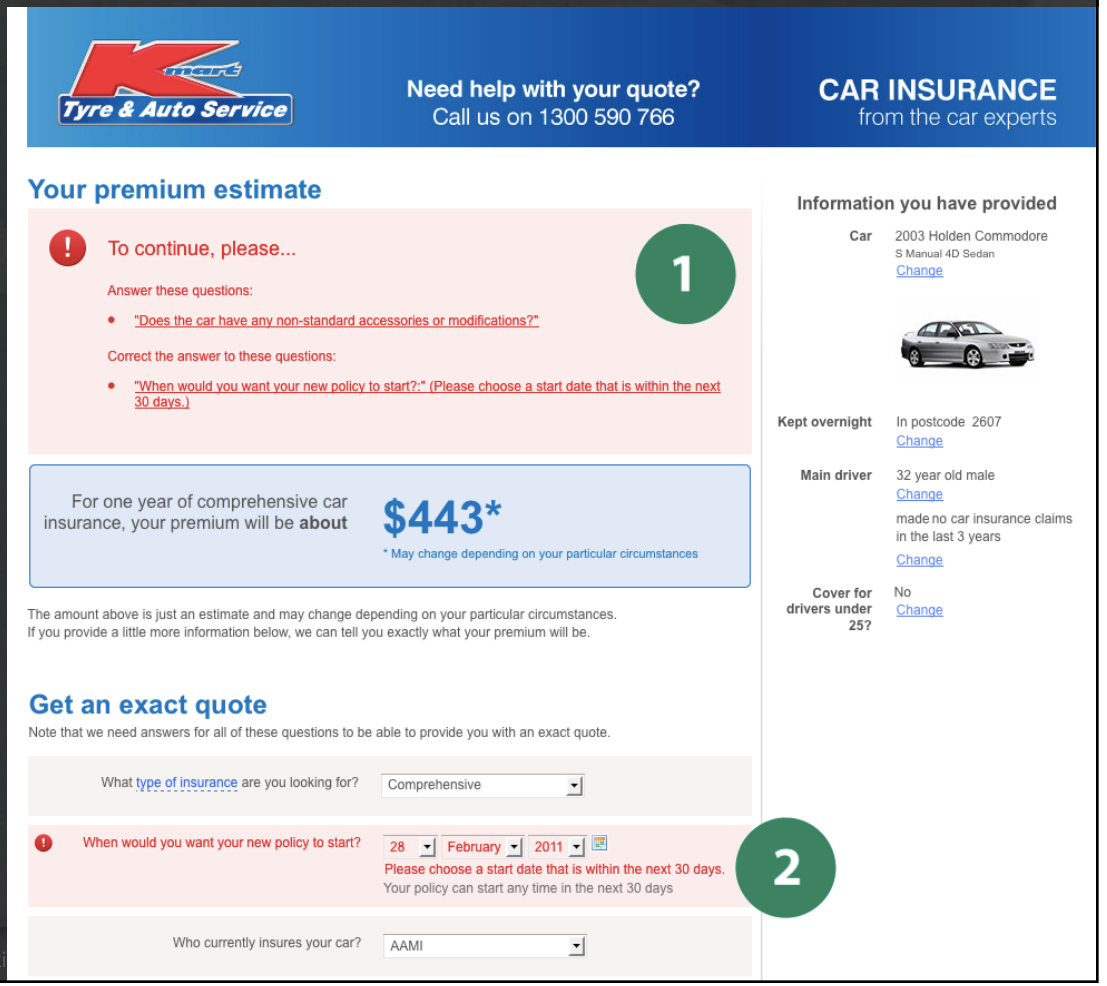
Heuristics (cont.)

good



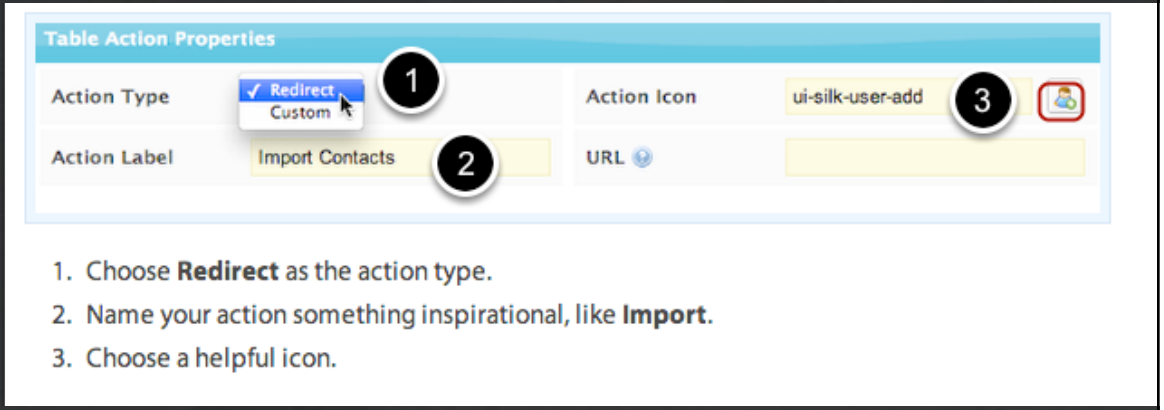
Good Error Messages

- Clearly indicate what has gone wrong
- Human readable
- Polite
- Describe the problem
- Explain how to fix it
- Highly noticeable



H10 – Help & Documentation

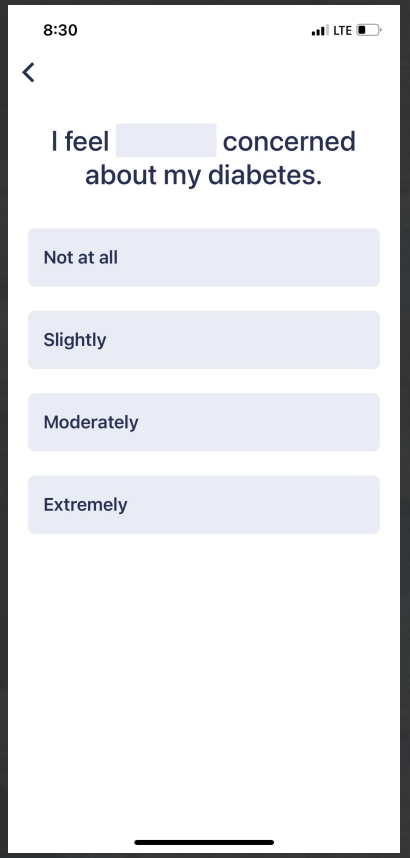
- Better if the system can be used without documentation, but it may be necessary
- How
  - easy to search
  - focused on task
  - list concrete steps



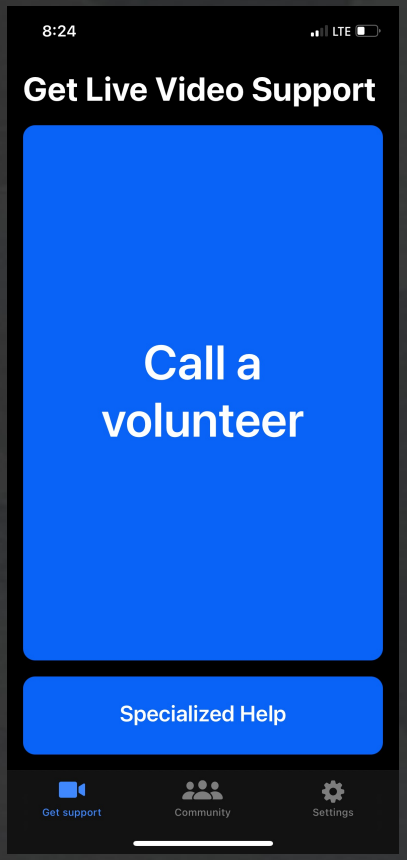
http://blog.screensteps.com/10-examples-of-great-end-user-documentation

H11* – Accessible

Users can interact with the system using *alternative input methods*. Content is legible with *distinguishable contrast* & *text size*. Key information is upfront & *not nested* for screen readers. Purely visual or auditory content has *text-based alternatives* for users with low vision & low hearing.



Glucose Buddy

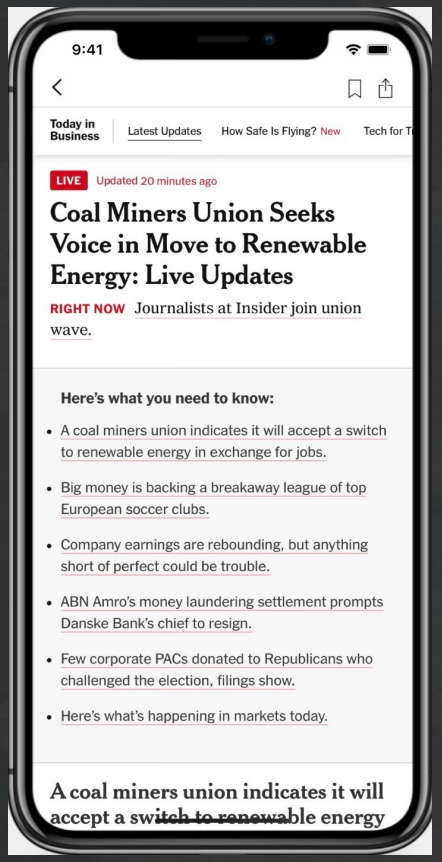


Be My Eyes

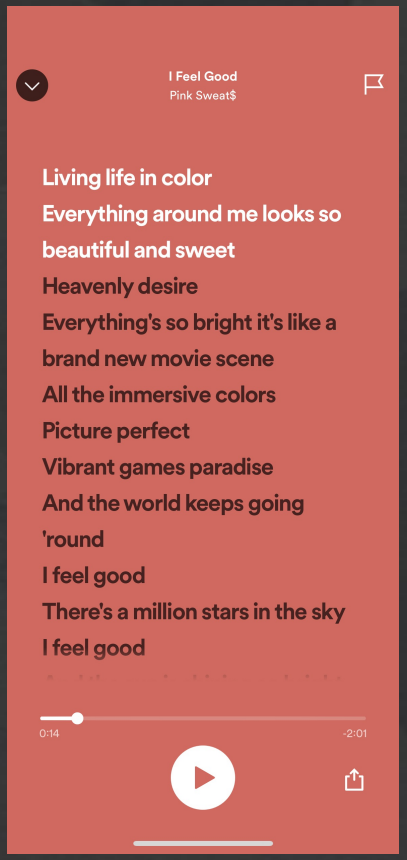
* New heuristic that CS147 staff has added to Nielson's 10.

H11* – Accessible

Users can interact with the system using *alternative input methods*. Content is legible with *distinguishable contrast* & *text size*. Key information is upfront & *not nested* for screen readers. Purely visual or auditory content has *text-based alternatives* for users with low vision & low hearing.



NY Times

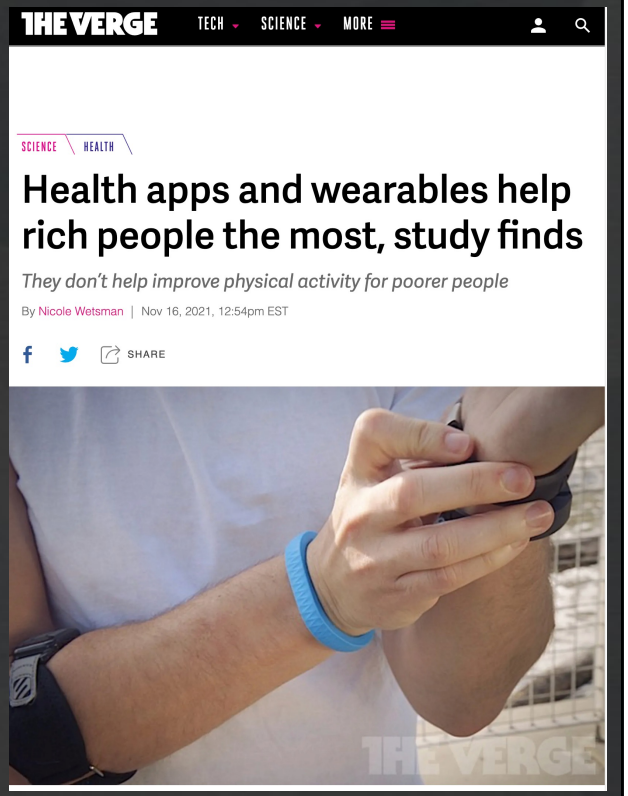


Spotify (lyrics mode)

* New heuristic that CS147 staff has added to Nielson's 10.

H12* – Fairness & inclusion

Users shouldn't feel that the design is not made for them. The design should meet all users' needs equally and *prevent the reproduction of pre-existing inequities*. It should *not create additional burdens for disadvantaged populations*.



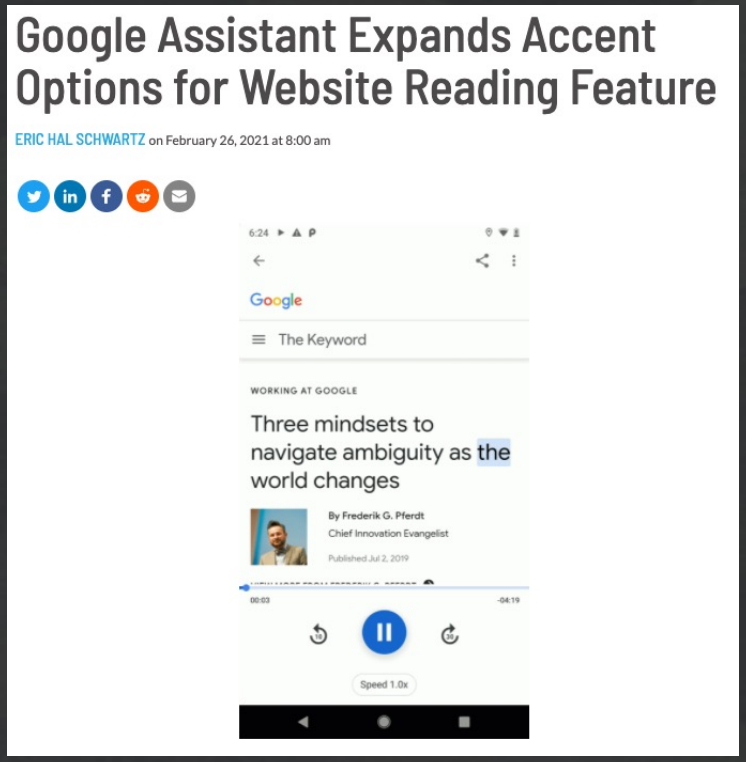
https://www.theverge.com/2021/11/16/22785163/digital-health-physical-activity-inequity

Western, M.J., Armstrong, M.E.G., Islam, I. *et al.* The effectiveness of digital interventions for increasing physical activity in individuals of low socioeconomic status: a systematic review and meta-analysis. *Int J Behav Nutr Phys Act* **18,** 148 (2021). https://doi.org/10.1186/s12966-021-01218-4

* New heuristic that CS147 staff has added to Nielson's 10.

H13* – Value alignment

The design should *encode values that users can understand & relate to*. *Conflicting collateral values should not emerge* when the user interacts with the product. Encoded values should match users' values in a broad set of contexts.



Google adds accented reading but leaves out marginalized groups.

* New heuristic that CS147 staff has added to Nielson's 10.

Heuristic Violation Examples

1. [H6 Recognition Rather Than Recall]
   Can't copy info from one window to another
   - user needs to memorize the data & retype
   - fix: allow copying

2. [H4 Consistency and Standards]
   Typography uses different fonts in 3 dialog boxes
   - slows users down
   - probably wouldn't be found by user testing
   - fix: pick a single format for entire interface

Severity Ratings

0 - don’t agree that this is a usability problem

1 - cosmetic problem

2 - minor usability problem

3 - major usability problem; important to fix

4 - usability catastrophe; imperative to fix

Severity Ratings Example

1. [H4 Consistency & Standards] [Severity 3]

The interface used the string “Save” on the first screen for saving the user’s settings, but used the string “Store” on the second screen. Users may be confused by this different terminology for the same function.

Fix: Use “Save” everywhere in the application.

Carl's Car Audio and Electronics -- Shopping Bag

what fits my car? logout continue shopping

Your vehicle: 1989 Tercel
To select a different vehicle, click 'What Fits My Car' above

PLEASE NOTE

Item #	Message
204EL570	This component is **NOT** recommended for your vehicle. We suggest removing this item (bolded in red below) from your cart. Please call us toll-free at 1-888-955-6000 and we'll be glad to provide further assistance.

Cart

Quantity	Item #	Remove?	Item Description	Price	Total
3	#033DVM4800	☐	Denon DVM-4800 DVD Video Changer	$1,199.95	$1,199.95
1	#204EL570	☐	Bazooka EL570 5"x7" 2-way speakers Vehicle: 1989 Toyota Tercel Liftback with equalizer or radio delete option (change)	$79.95	$79.95
h	#158DSCP50	☐	**Special** Sony DSC-P50 Digital Still Camera, 2.1 Mega Pixel Save $100! Was $399.95, Now: $299.95	$299.95	$299.95
1	#133DVDLA95	yes ☐ no ☐	Panasonic DVD-LA95 Portable DVD-A/V Player with 9" Screen -- This item is temporarily out of stock --	$999.95	$999.95
1	#170GXT160	☐	Maxell GX-Silver T-160 VHS Video Tape	$2.49	$2.49
1					

Total

Merchandise Total:	**$2,582.29**
Standard Shipping Shipping Charge: Alternative shipping options available before final checkout	**$13.95**
Order Total:	**$2,596.24**

Update Cart Clear Cart

- To change an item's quantity, enter the correct number in the **Quantity** column, then press **Update Cart**.
- To remove an item, check the box in the **Remove?** column, then press **Update Cart**.
- To order an item that appears in your printed Crutchfield catalog, enter the item number into the Cart and click **Update Cart**.
- International visitors, please click here.

Problems Found this Year

- H1: Visibility of system status
  - Username for current shopping cart missing [1]
- H2: Match between system & real world
  - "price" and "total" should refer to different things [1]
  - Items referred to by their Item Number [3]
- H3: User control & freedom
  - Not clear where/how user can place order [1]
  - Missing a search bar [1]
- H4: Consistency & standards
  - Red used inconsistently - error, special, title, out of stock [4]
  - Yes/No Checkbox in “Remove?’ [5]
  - Menu Bar links on different lines - logout, continue, shopping [1]

Problems Found this Year

H5: Error prevention

- Out of Stock items allowed to be added to cart [4]
- ”h” in Quantity Field [7]

- H6: Recognition rather than recall
  - Item number not auto-filled, difficult for user to remember [2]
- H7: Flexibility and efficiency of use
  - Update Cart Button required to update cart and ambiguous use [2]
  - Quantity Field could be a selection menu [1]
- H8: Aesthetic & minimalist design
  - Color Combinations - red/green, black text on blue background [6]
  - Crowded Design - columns and text squished [2]
  - # Sign not necessary [1]

Problems Found this Year

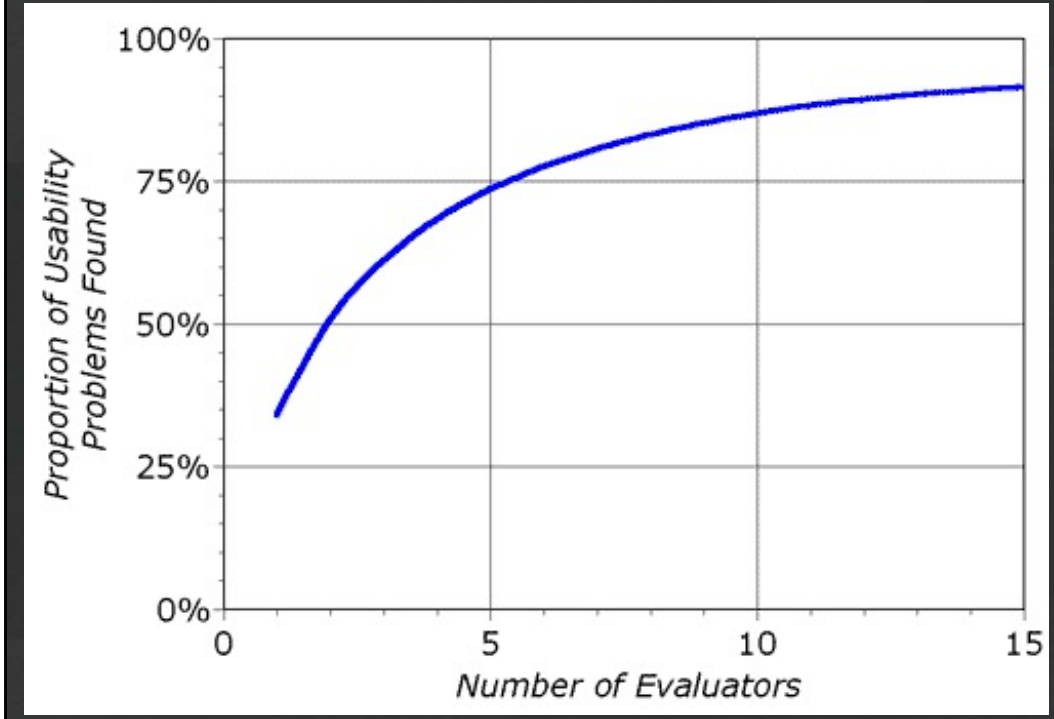
- H9: Help users recognize, diagnose, & recover from errors
  - “Please Note” error message refers to item bolded in red but multiple items bolded in red [2]
  - Last row has quantity 1 but no other info, user unsure how to fix [1]
- H10: Help & Documentation
  - Instructions for use not immediately visible [1]
  - International users not immediately redirected to other page [2]
- H11*: Accessible
  - Aesthetic choices difficult for visually impaired- blue links, small text, blue on blue color scheme [3]
  - Error message relies on ability to view red color - hard for color-blind [1]
- H12*: Fairness & inclusion
  - Requires knowledge of your car beyond just model and year [1]

Problems Found Last Year

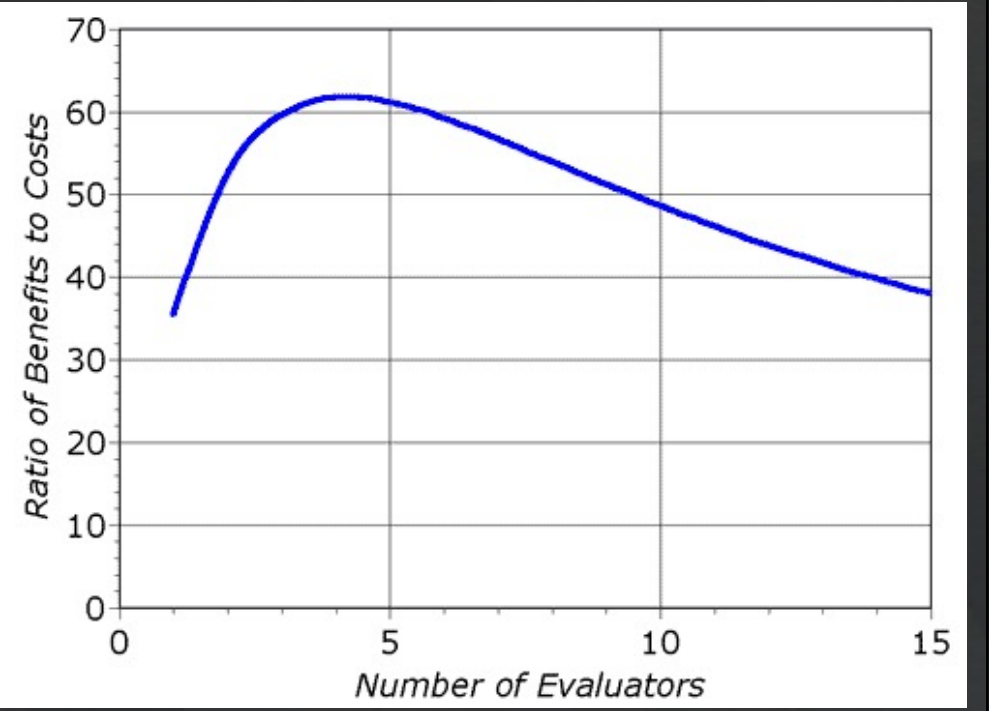
1. H4 (consistency): The use of red to indicate errors, out of stock items, and sales/savings [26]
2. H4 (consistency): two of the check boxes have yes/no next to them but none of the others do. Error prevention? [22]
3. H9 (Aesthetic): "#" is not needed [2]
4. H5 [Error Prevention]: “h” in the quantity box. [11]

Decreasing Returns

problems found



benefits / cost



* Caveat: graphs for a specific example

Heuristic Evaluation Summary

- Have evaluators go through the UI twice
- Ask them to see if it complies with heuristics
  - note where it doesn't & say why
- Have evaluators independently rate severity
- Combine the findings from 3 to 5 evaluators
  - come to agreement on problems, fixes & severity

- Alternate with user testing

Speech UI Heuristics

S1: Give the agent a persona through language, sounds, and other styles.

S2: Make the system status clear.

S3: Speak the user's language.

S4: Start and stop conversations.

S5: Pay attention to what the user said and respect the user's context.

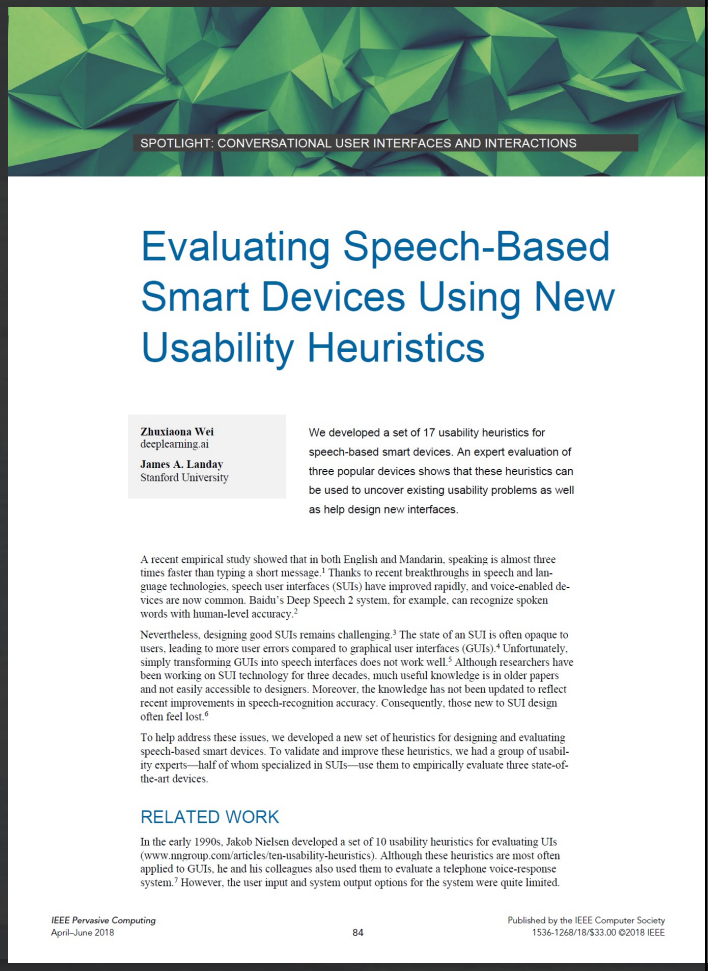
S6: Use spoken language characteristics.

S7: Make conversation a back-and-forth exchange.

S8: Adapt agent style to who users are, how they speak, and how they are feeling.

S9: Guide users through a conversation so they are not easily lost.

S10: Use responses to help users discover what is possible.



SPOTLIGHT: CONVERSATIONAL USER INTERFACES AND INTERACTIONS

Evaluating Speech-Based Smart Devices Using New Usability Heuristics

Zhuxiaona Wei
deeplearning.ai

James A. Landay
Stanford University

We developed a set of 17 usability heuristics for speech-based smart devices. An expert evaluation of three popular devices shows that these heuristics can be used to uncover existing usability problems as well as help design new interfaces.

A recent empirical study showed that in both English and Mandarin, speaking is almost three times faster than typing a short message.[1] Thanks to recent breakthroughs in speech and language technologies, speech user interfaces (SUIs) have improved rapidly, and voice-enabled devices are now common. Baidu's Deep Speech 2 system, for example, can recognize spoken words with human-level accuracy.[2]

Nevertheless, designing good SUIs remains challenging.[3] The state of an SUI is often opaque to users, leading to more user errors compared to graphical user interfaces (GUIs).[4] Unfortunately, simply transforming GUIs into speech interfaces does not work well.[5] Although researchers have been working on SUI technology for three decades, much useful knowledge is in older papers and not easily accessible to designers. Moreover, the knowledge has not been updated to reflect recent improvements in speech-recognition accuracy. Consequently, those new to SUI design often feel lost.[6]

To help address these issues, we developed a new set of heuristics for designing and evaluating speech-based smart devices. To validate and improve these heuristics, we had a group of usability experts—half of whom specialized in SUIs—use them to empirically evaluate three state-of-the-art devices.

RELATED WORK

In the early 1990s, Jakob Nielsen developed a set of 10 usability heuristics for evaluating UIs (www.nngroup.com/articles/ten-usability-heuristics). Although these heuristics are most often applied to GUIs, he and his colleagues also used them to evaluate a telephone voice-response system.[7] However, the user input and system output options for the system were quite limited.

IEEE Pervasive Computing
April–June 2018

Published by the IEEE Computer Society
1536-1268/18/$33.00 ©2018 IEEE

Speech UI Heuristics

S11: Keep feedback and prompts short.

S12: Confirm input intelligently.

S13: Use speech-recognition system confidence to drive feedback style.

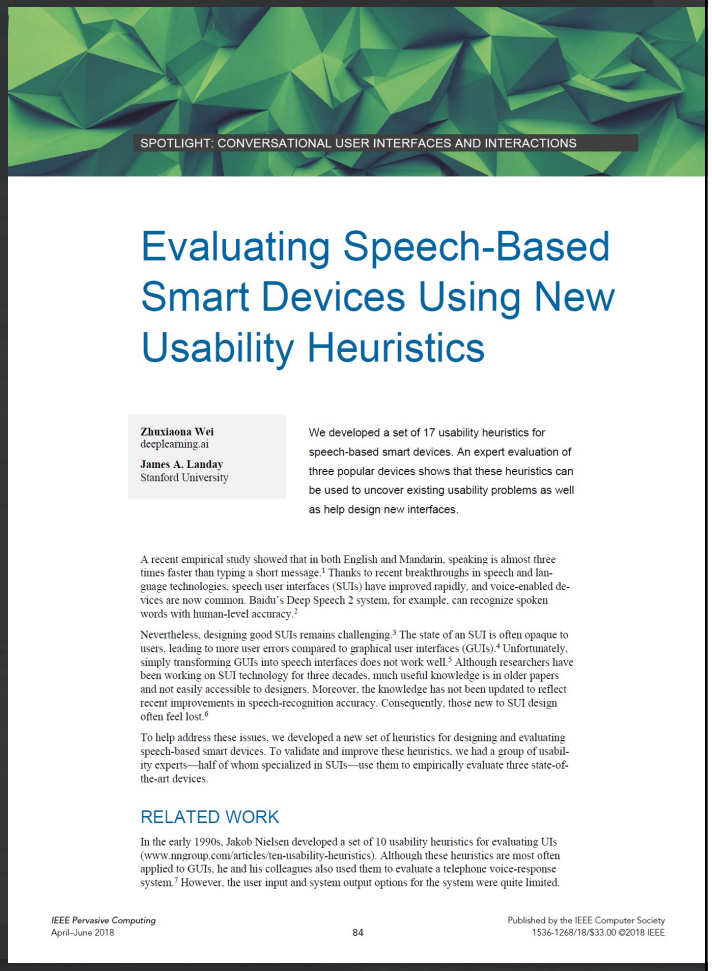
S14: Use multimodal feedback when available.

S15: Avoid cascading correction errors.

S16: Use normal language in communicating errors.

S17: Allow users to exit from errors or a mistaken conversation.

The list of heuristics along with detailed descriptions and examples can be found at http://hci.stanford.edu/publications/2018/speech-he/sui-heuristics.html



SPOTLIGHT: CONVERSATIONAL USER INTERFACES AND INTERACTIONS

Evaluating Speech-Based Smart Devices Using New Usability Heuristics

Zhuxiaona Wei
deeplearning.ai

James A. Landay
Stanford University

We developed a set of 17 usability heuristics for speech-based smart devices. An expert evaluation of three popular devices shows that these heuristics can be used to uncover existing usability problems as well as help design new interfaces.

A recent empirical study showed that in both English and Mandarin, speaking is almost three times faster than typing a short message.[1] Thanks to recent breakthroughs in speech and language technologies, speech user interfaces (SUIs) have improved rapidly, and voice-enabled devices are now common. Baidu's Deep Speech 2 system, for example, can recognize spoken words with human-level accuracy.[2]

Nevertheless, designing good SUIs remains challenging.[3] The state of an SUI is often opaque to users, leading to more user errors compared to graphical user interfaces (GUIs).[4] Unfortunately, simply transforming GUIs into speech interfaces does not work well.[5] Although researchers have been working on SUI technology for three decades, much useful knowledge is in older papers and not easily accessible to designers. Moreover, the knowledge has not been updated to reflect recent improvements in speech-recognition accuracy. Consequently, those new to SUI design often feel lost.[6]

To help address these issues, we developed a new set of heuristics for designing and evaluating speech-based smart devices. To validate and improve these heuristics, we had a group of usability experts—half of whom specialized in SUIs—use them to empirically evaluate three state-of-the-art devices.

RELATED WORK

In the early 1990s, Jakob Nielsen developed a set of 10 usability heuristics for evaluating UIs (www.nngroup.com/articles/ten-usability-heuristics). Although these heuristics are most often applied to GUIs, he and his colleagues also used them to evaluate a telephone voice-response system.[7] However, the user input and system output options for the system were quite limited.

IEEE Pervasive Computing
April–June 2018

Published by the IEEE Computer Society
1536-1268/18/$33.00 ©2018 IEEE

Administrivia

- Questions on individual heuristic evaluation assignment?

Grading on Last Assignment

Assignment #5 Low-fi Prototype & Test

A5 Group Report:	✓--: 0%	✓-: 18%	✓: 71%	✓+: 11%	✓++: 0%
A5 Group Presentation:	✓--: 0%	✓-: 7%	✓: 64%	✓+: 29%	✓++: 0%
A5 Individual Presentation:	✓--: 0%	✓-: 2%	✓: 22%	✓+: 73%	✓++: 4%

A5 Group Report Average: 90%

A5 Group Presentation Average: 91%

A5 Individual Presentation Average: 94%

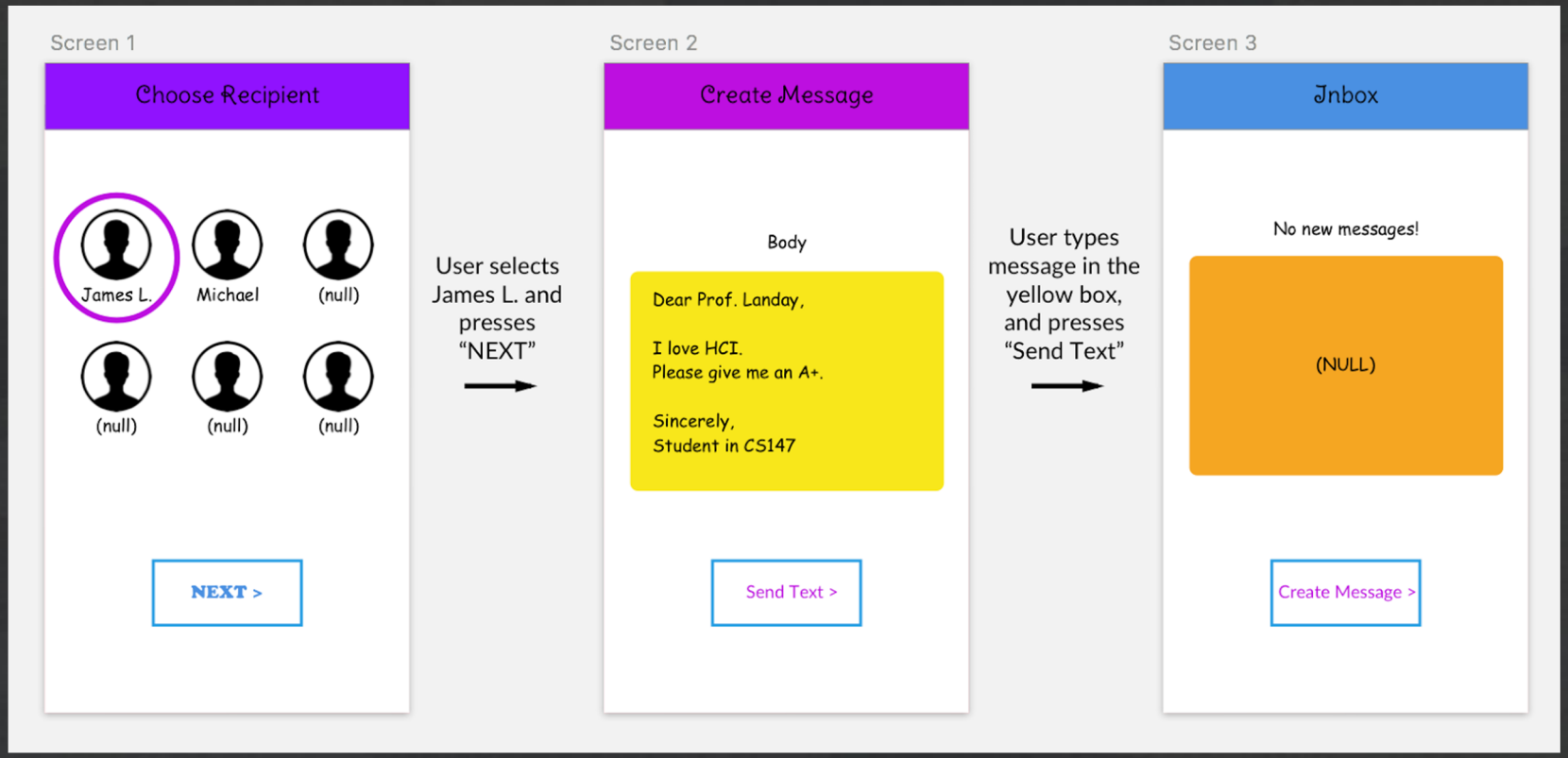
TEAM BREAK

EXERCISE

Heuristic evaluation practice

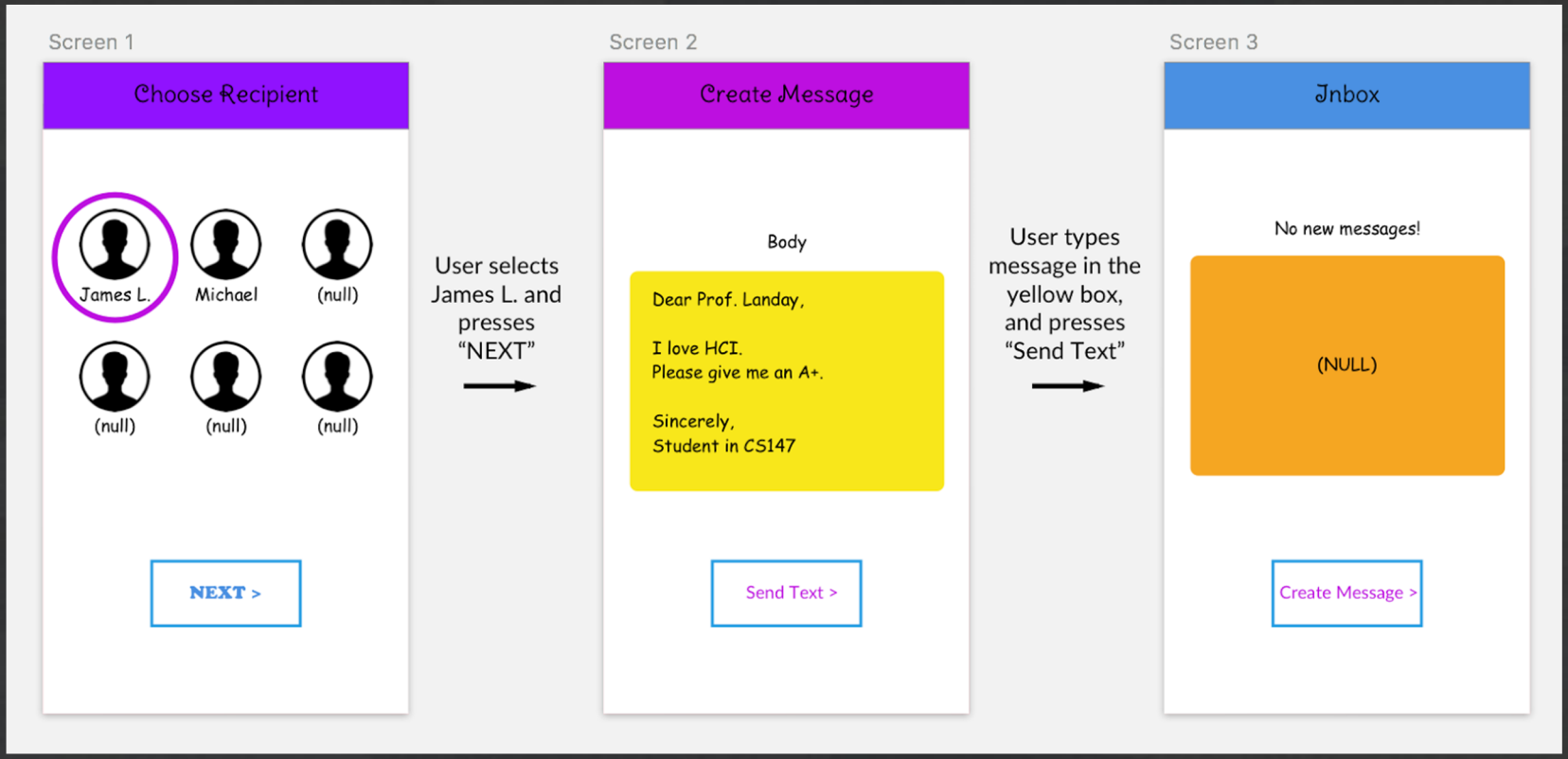
[8 min]: On your own, find 8+ usability violations

For each one: Which guideline was violated and why? How could it be fixed?



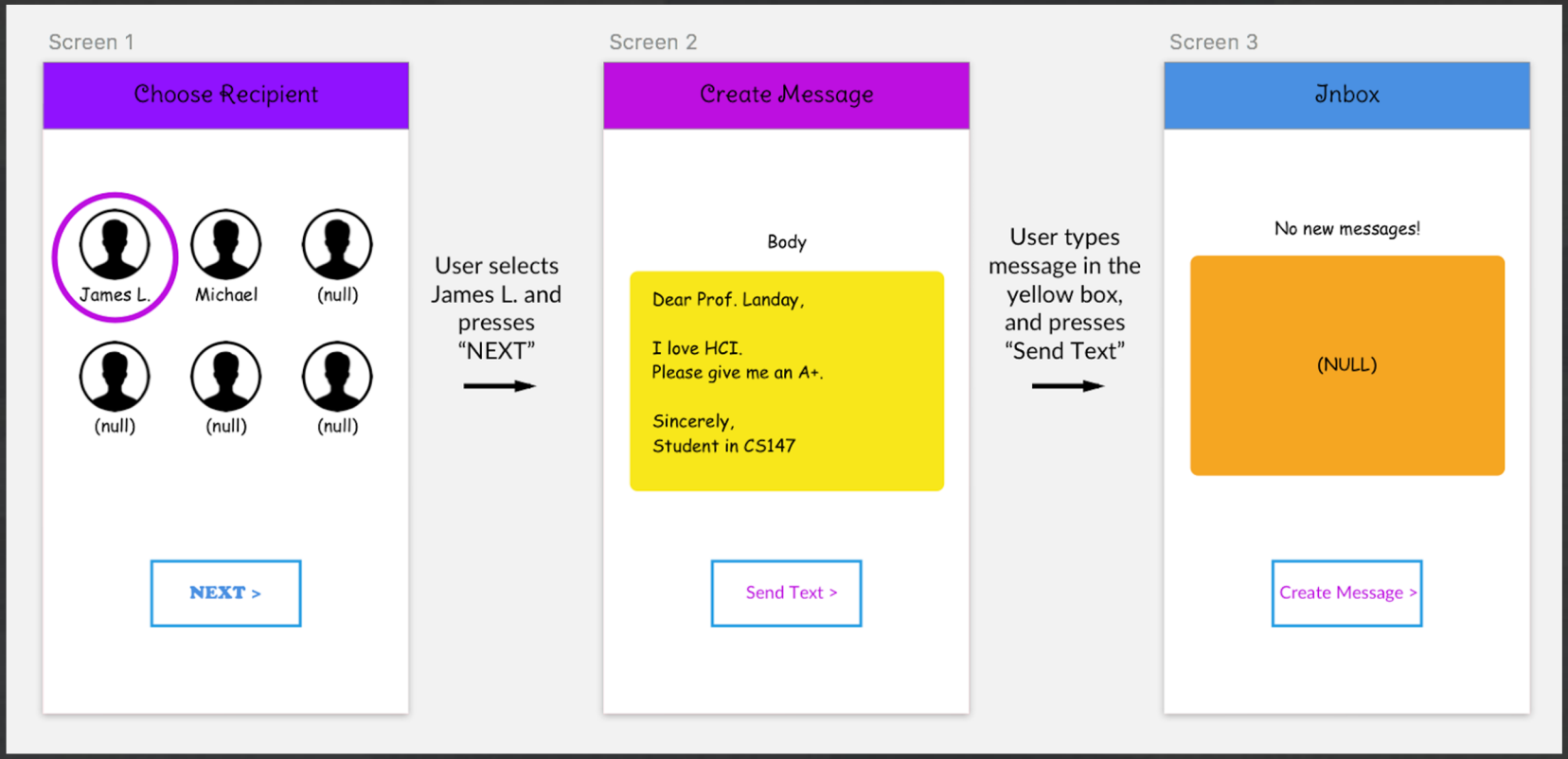
[8 min]: Share with your breakout room

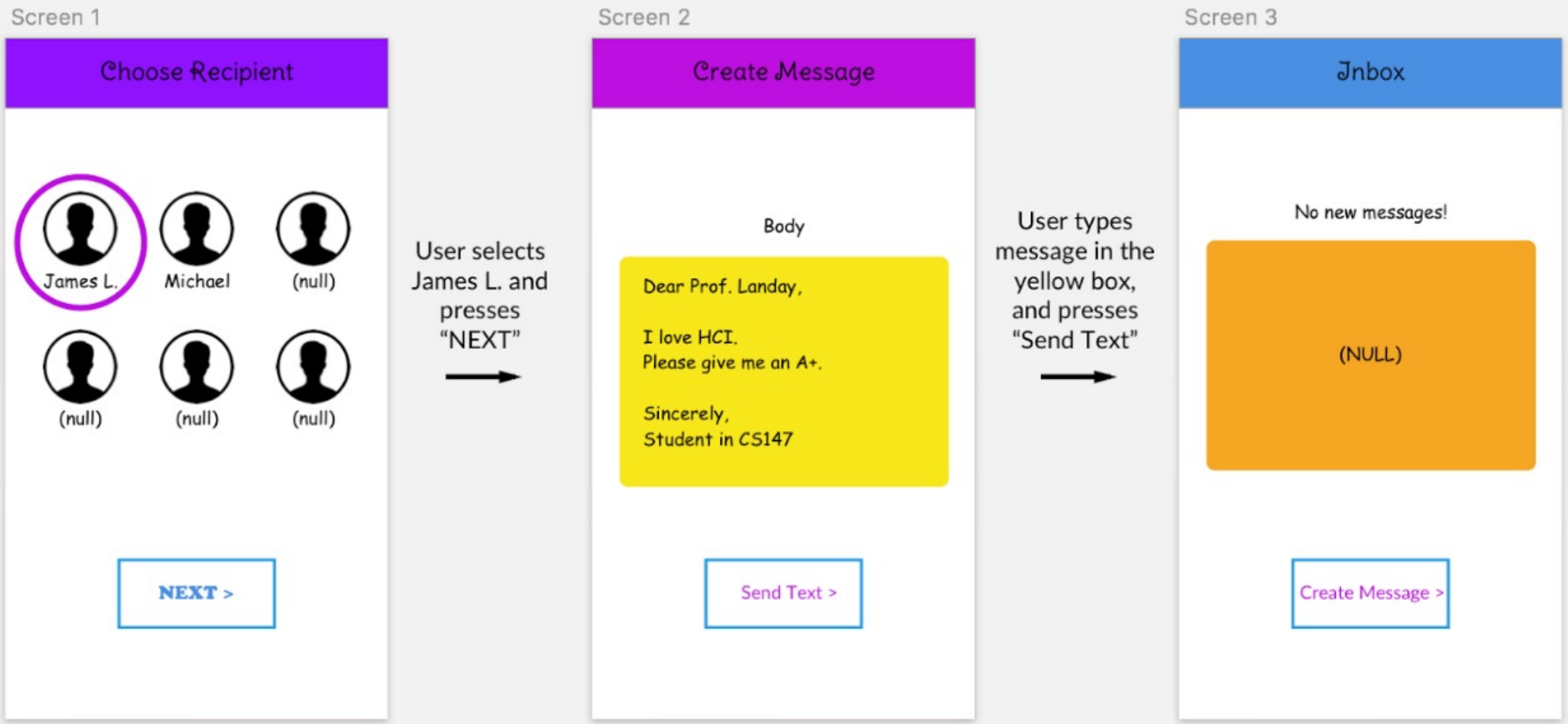
What violations did you all find? Which ones did only one of you notice?



[10 min]: Share with the class

Raise your hand or post a screenshot with your description on Slack.

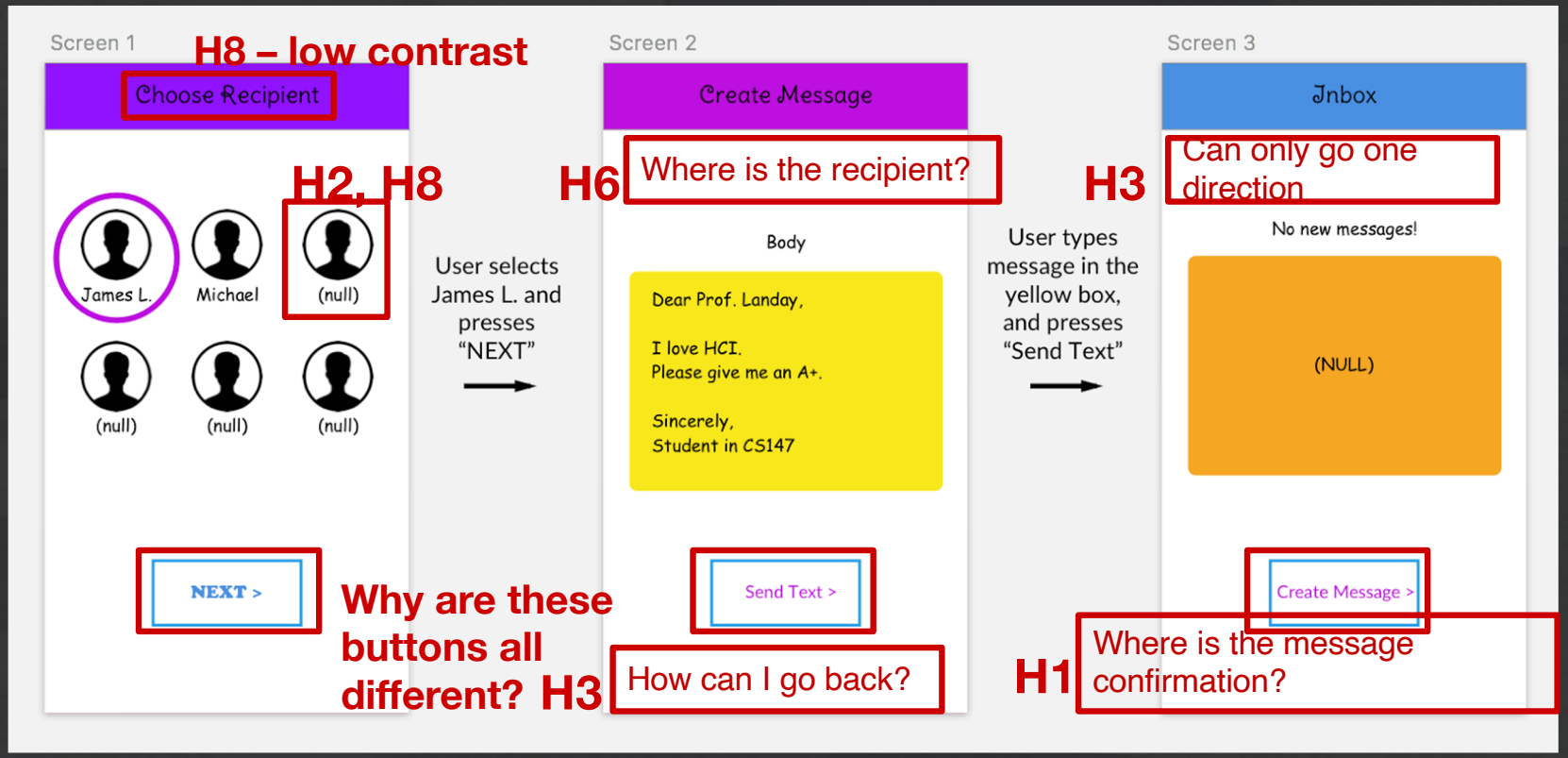




Find, label, & describe 8-10 Heuristic Violations

[10 min]: Share with the class

Raise your hand or post a screenshot with your description on Slack.



Problems Found This Year

H2: Match between system & real world

Showing “(null)” for empty content

H3: User control & freedom

No “back” button [5]

H6: Recognition rather than recall

No way to see who you’re sending to in the second screen

Problems Found Last Year

1. H1: no feedback on whether the message is sent or not, just navigates to inbox screen
2. [H3: user control and freedom] No way to cancel message solution: add an ‘x’ or cancel button that allows them to delete the message on screen 2
3. H8. There was unnecessary users listed that were null
4. H4: so many different colors, not consistent
5. H1: No visibility of system status (can’t really tell which page I’m on - perhaps some sort of diagram at the bottom could fix this)
6. H6: James L is not displayed on create message screen
7. [H6 Recognition rather than recall] Can’t see who you’re sending the message to when you’re creating the message. Fix: include sub-header that specifies name you just clicked on
8. H4: inconsistent button design
9. H4: consistency with names on the first screen (names have or don’t have last initials) --> should just choose one format (potentially just first name if we want to avoid clutter)
10. H8 (aesthetics) - The headers have low contrast (especially the purple/pink) and should use less saturated colors for better contrast.
11. H3/H6: User can’t go back from screen 2 to 1 if they miscliked on the wrong person’s profile, no recall of who you selected --> recall the recipient name on screen 2
12. H3: User control and freedom - can only send to recipients in the list, no way to add a new recipient
13. H9: “error”/null messages offer no explanation
14. H3: can only send text (no images/files which might be helpful in certain cases)
15. [H5 Error prevention] Could potentially select (null) users to write messages to. Fix: just don’t display the icons of null users at all.
16. H7: unnecessary “next” button on the first page. Might instead click on the person to transition to the next screen, and display the person’s name on screen 2 (H6) and add back button there (H3)
17. H3: How do you view old messages?
18. H10: no help or documentation included - just buttons indicating the very next step
19. H4 consistency: not sure the differences between ‘text’ ‘messages’ ‘body’
20. H3 - no ability to edit message text
21. H2: "Body" and "Null" might not make sense to user without technical background, but also why would they even see these words to begin with
22. H4: terminology of "message" vs "text"

Further Reading

Heuristic Evaluation

- Longer lecture
  - https://drive.google.com/file/d/0BweiB6wu4sBaN2tfZGxKb2tuOTg/view
- Books
  - *Usability Engineering*, by Nielsen, 1994
- Web site
  - http://www.nngroup.com/articles/

Next Time

- Conceptual Models & Interface Metaphors
  - Read "The Psychology of Everyday Things" (Ch. 1), from *The Design of Everyday Things* by Donald Norman
- Midterm Review
  - Thur evening (6:30-7:30 PM) on Zoom (class zoom)
- Studio
  - Ad-hoc group heuristic evaluation
  - Must be present to get credit on assignment