

**Building User Interfaces**

***Design Principles for***

**Voice User Interfaces**

**Professor Yuhang Zhao**

# What will we learn today?

- Design principles for conversational interfaces
- Usability heuristics for conversational interfaces

# Design Principles for Conversational Interfaces

## Elephant in the Room

**Recap: Definition of Usability:** The effectiveness, efficiency, and satisfaction with which a specified set of users can achieve a specified set of tasks in a particular environment. — ISO 9241-11

Conversational interfaces are almost always less effective, less efficient, and less satisfactory than graphical user interfaces.

**Effectiveness:** Conversational interfaces are more error prone due to technology, ambiguities, and environmental influences.

**Efficiency:** Using conversational interfaces is almost never as fast as using graphical user interfaces.

**Satisfaction:** Interacting with conversational interfaces can be awkward, socially inappropriate, and frustrating.

So, what is the point of conversational interfaces?

## Where do these interfaces deliver value?

1. Streamlining app installation, login, payment, notifications, and and so on in a conversational paradigm.<sup>1</sup>
2. In some contexts, e.g., while driving, CIs are more effective, efficient, and satisfactory due to resource constraints.
3. CIs address many accessibility problems, including vision (e.g., blindness), motor (e.g., tremor), and cognitive (e.g, dyslexia) deficiencies.

<sup>1</sup> Grover, 2016, Bots won't replace apps. Better apps will replace apps.

# Design Principles

## Gricean Maxims<sup>2</sup>

**Definition:** Proposed by Paul Grice, conversations follow the cooperative principle and four key maxims:

- Maxim of quality (truthful and accurate communication)
- Maxim of quantity (just the right amount of information)
- Maxim of relevance (appropriate and relevant information)
- Maxim of manner (clear, cooperative communication)

<sup>2</sup> Grice, 1975, *Logic and Conversation*

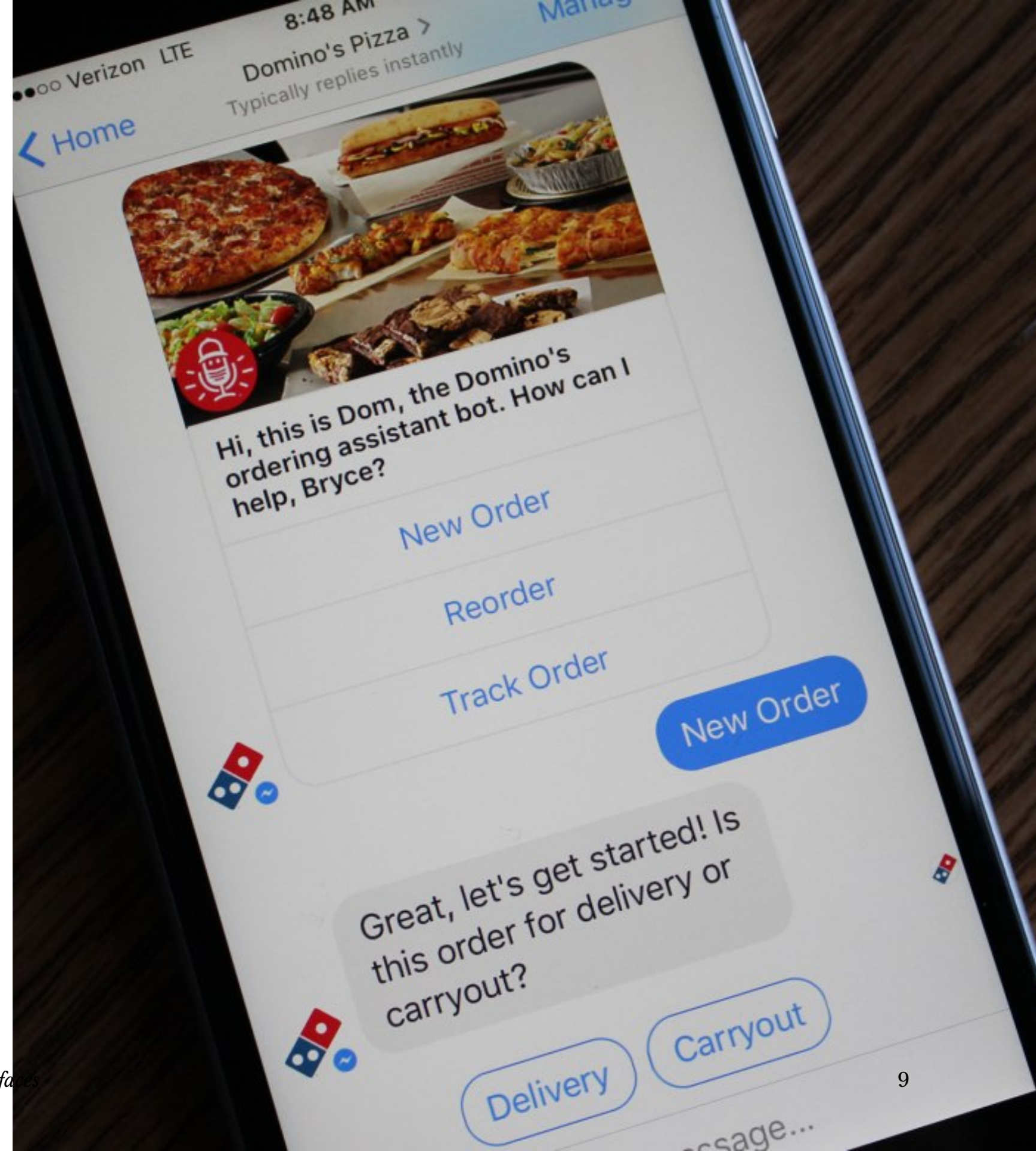


# Multimodality<sup>3</sup>

**Definition:** Multimodal interfaces utilize multiple modalities, including visual information, speech, touch, and so on, in user experiences they afford.

Most conversational interfaces are multimodal interfaces.

<sup>3</sup>Image source



**Multimodality Principle:** Take advantage of other modalities, e.g., visual information, vibrations, etc., wherever appropriate.

Using multimodal components, you can provide users with breaks for decision making, interruptions, etc.

Potential caveats:

- Ask, "does my interface still support a speech-only interaction?"
- The conversational and other components must be designed together to fit within the conversation.

## Interaction Paradigm

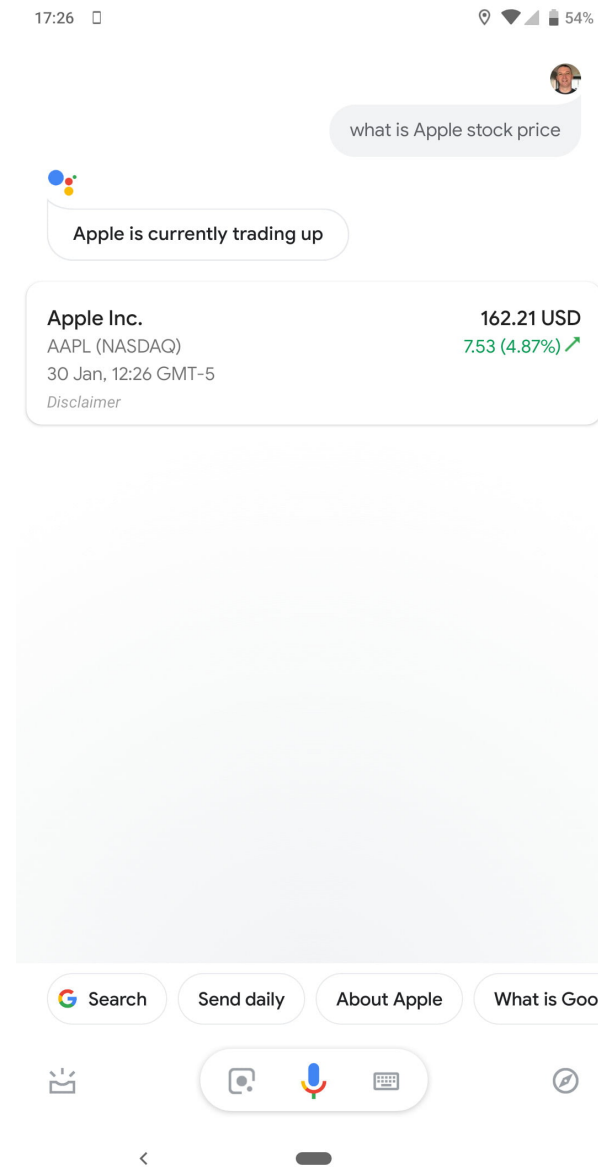
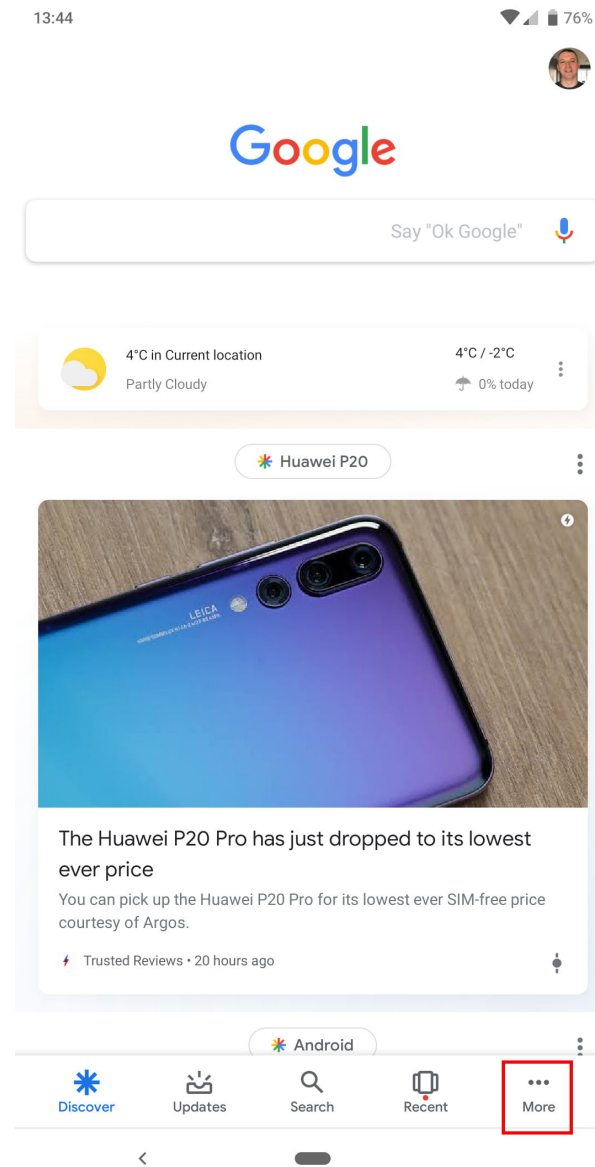
Conversational interfaces can follow different paradigms depending on the context of use and the design of the application:

- Command-and-control interfaces (always-on voice assistants)
- Conversational interfaces (chatbots, task assistants, social robots)

## Command-and-Control Interfaces

**Definition:** Interfaces where speech input is mapped to specific system functions that are called immediately. These interfaces commonly utilize:

1. Expressing user intent using a wake word (e.g., "OK, Google") or the pressing of a button (e.g., home button in the iPhone)
2. Indicating listening and understanding
3. Executing the mapped function



<sup>4</sup> Image source: left, right, middle

# Conversational Interfaces

**Definition:** Interfaces where the interaction with the system has the characteristics of human conversations, including turn taking, theoretically infinite depth, conversational markers, etc.

## Turn-taking

**Definition:** Speaking turns are the core, cooperative structure of conversations that involves one speaker at a time and an explicit exchange of tokens.

### Principles:

1. One speaker at a time — *transparency in who is speaking*
2. Turn exchanges — *explicit signaling of who will speak next*
3. Interruption handling — *very difficult with CIs*

## Conversational Markers<sup>5</sup>

**Definition:** Speech cues that indicate the state or the direction of the conversation. Types of conversational markers:

1. Timelines ("First," "Halfway there," "Finally")
2. Acknowledgements ("Thanks," "Got it," "Alright," "Sorry about that")
3. Positive feedback ("Good job," "Nice to hear that")

<sup>5</sup> Pearl, 2016, *Designing Voice User Interfaces: Principles of Conversational Experiences*



**Assistant:** I'll be asking you a few questions about your health. First, how many hours of sleep did you get last night?

**User:** About seven.

**Assistant:** Good job. And how many serving of fruits and vegetables did you eat yesterday?

**User:** Maybe four.<sup>5</sup>

<sup>5</sup> Pearl, 2016, *Designing Voice User Interfaces: Principles of Conversational Experiences*

**Assistant:** Got it. Last question—were you able to take your medication last night?

**User:** Yes.

**Assistant:** All right. That's it for now. I'll talk to you again tomorrow. Goodbye.<sup>5</sup>

<sup>5</sup> Pearl, 2016, *Designing Voice User Interfaces: Principles of Conversational Experiences*

## Confirmations<sup>5</sup>

**Definitions:** CIs are designed with explicit forms of confirmation to improve system usability and transparency.

Can be explicit vs. implicit and speech-based vs. non-speech based (visual, action).

<sup>5</sup> Pearl, 2016, *Designing Voice User Interfaces: Principles of Conversational Experiences*

**Explicit confirmation:** Requiring the user to confirm: "I think you want to set a reminder to 'buy insurance before going skydiving next week.' Is that right?"<sup>5</sup>

**Implicit confirmation:** Letting user know what was understood: "Ok, setting a reminder to buy insurance..."

<sup>5</sup> Pearl, 2016, Designing Voice User Interfaces: Principles of Conversational Experiences

## Error Handling

**Definitions:** Deviations from expected conversational flow due to technical mistakes, unexpected user behavior, environmental influences, etc.

## Types of errors:<sup>5</sup>

1. No speech detected
2. Speech detected, but nothing recognized
3. Something was recognized correctly, but the system does the wrong thing with it
4. Something was recognized incorrectly

<sup>5</sup> Pearl, 2016, *Designing Voice User Interfaces: Principles of Conversational Experiences*



# Usability Heuristics for Conversational Interfaces



## Recap: What are Usability Heuristics?<sup>6</sup>

**Definition:** Developed by Jacob Nielsen, heuristic evaluation involves having a small set of evaluators examine the interface and judge its compliance with recognized usability principles (the "heuristics").

<sup>6</sup> NN/g: How to conduct a heuristic evaluation

# Heuristics for Conversational Interfaces<sup>7</sup>

Seventeen heuristics that fall into five broad categories:

1. General
2. Conversational style
3. Guiding, Teaching, and Offering Help
4. Feedback and prompts
5. Errors

<sup>7</sup>Wei & Landay, 2018, [Speech-based Conversational Agent Heuristics](#)

# General Heuristics

## Heuristic #1

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

- Create an illusion by being consistent.
- Make sure to do this without being distracting.

## Heuristic #2

### **S2: Make the system status clear.**

- Use verbal, sound, or multimodal feedback.
- Communicate delays immediately and give feedback while “busy”.

## Heuristic #3

### S3: Speak the user's language.

- Use words, phrases and concepts familiar to end users, rather than system-oriented or technical jargon.

## Heuristic #4

### S4: Start and stop conversations.

- Use a wake word to start a conversation, but don't require it again in the same conversation.
- Gracefully end conversations when the user is done.

Example: If the user doesn't speak for a 5-10 seconds, end the conversation and give feedback (e.g., a distinctive tone or lights) to indicate the conversation is done.

## Heuristic #5

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

— Leverage user input when it can be used as a parameter to a command.

**User: “Find a flight to San Francisco on June 25th”**

**Agent: “Searching for flights to San Francisco on June 25th...”**

— Remember what the user has said in the current conversation.

**User: “What is the weather like in San Jose today?”**

**User: “How about tomorrow?”**

**Agent: “It will be sunny in San Jose tomorrow.”**



- Use context you already know about the user to fill in fields, but confirm them.

**User: “Make a reservation for 2 tonight at Jack’s at 8pm”**

**Agent: “Ok, I made a reservation for John Smith for 2 people at Jack’s tonight at 8.”**

- Use context to respond intelligently (e.g., location/environment, time constraints, # of users, user identity/age).

**User: “What will the weather be like today?”**

**Agent: “The weather in Palo Alto will be sunny today. The high will be 77.”**

# Conversational Style

## Heuristic #6

### S6: Use spoken language characteristics.

- Use discourse markers as part of confirmations and prompts to make conversation more natural (e.g., “next”, “and”, “so”, “actually”, “sure”, “ok”, “got it”).
- Leverage prosody, including rhythm, tone, pauses, emphasis, discourse fillers (e.g., “uh”, “uhm”, “hmm”, “ah”, “like”).

## Heuristic #7

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

- Don't always prompt for everything all at once.
- Take turns and don't let instructions get in the way.
- Give users a chance before jumping in.

## Heuristic #8

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

- Users prefer agents that have conversational style similar to their own — match it.
- Match the user's emotion, gender, and personality.

# Guiding, Teaching, and Offering Help

## Heuristic #9

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

— Guide subtly using natural affordances rather than explicitly.

**User: “Check my email.”**

**Agent: “You have two new emails. The first email is ...”**

**Agent: “Repeat, reply or go to the next one?”**

- Guide user towards desired response and cue the user what type of response is desired.

**User: “Trivia Game”**

**Agent: “Choose from the following: capitals of the world, sports stars, or inventors”**

- Allow data to be naturally given in response to single or multiple prompts.

**User: “Get a reservation at Ding Tai Fang for tonight.”**

**Agent: “For what time at Ding Tai Fang?”**

**User: “7PM.”**

**Agent: “How many people tonight at 7?”**

**User: “Three.”**

**Agent: “Ok, I made a reservation for John Smith for 3 people at Ding Tai Fang tonight at 7.”**



## Heuristic #10

**S10: Use responses as a way to help users discover what is possible.**

- Teach multiple possible ways of asking for a result.
- Use examples in a natural manner rather than teaching commands explicitly.

# Feedback and Prompts

## Heuristic #11

### S11: Keep feedback and prompts short.

- Clear but succinct.
- Keep lists of items short (3-5 max.), and let people ask if they want to hear more.
- Let experienced users have faster and shorter prompts.

## Heuristic #12

### S12: Confirm input intelligently.

- Confirm input implicitly through results or next prompt.
- Confirm irreversible or critical actions explicitly and even allow undo after confirmation.

## Heuristic #13

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

- High: Do it and tell me
- Moderate: Confirm input
- Low: Re-prompt (“Say that again?”)

## Heuristic #14

### **S14: Use multimodal feedback when available.**

- Lights
- Graphic displays
- Sounds

# Errors

## Heuristic #15

### **S15: Avoid cascading correction errors.**

- Escalate detail in prompts when input is ambiguous or incorrect.
- If input results in multiple hypotheses, let user select from list with “yes” / “no”.
- For error correction, use a different modality or voice response style (e.g., select from a list).



## Heuristic #16

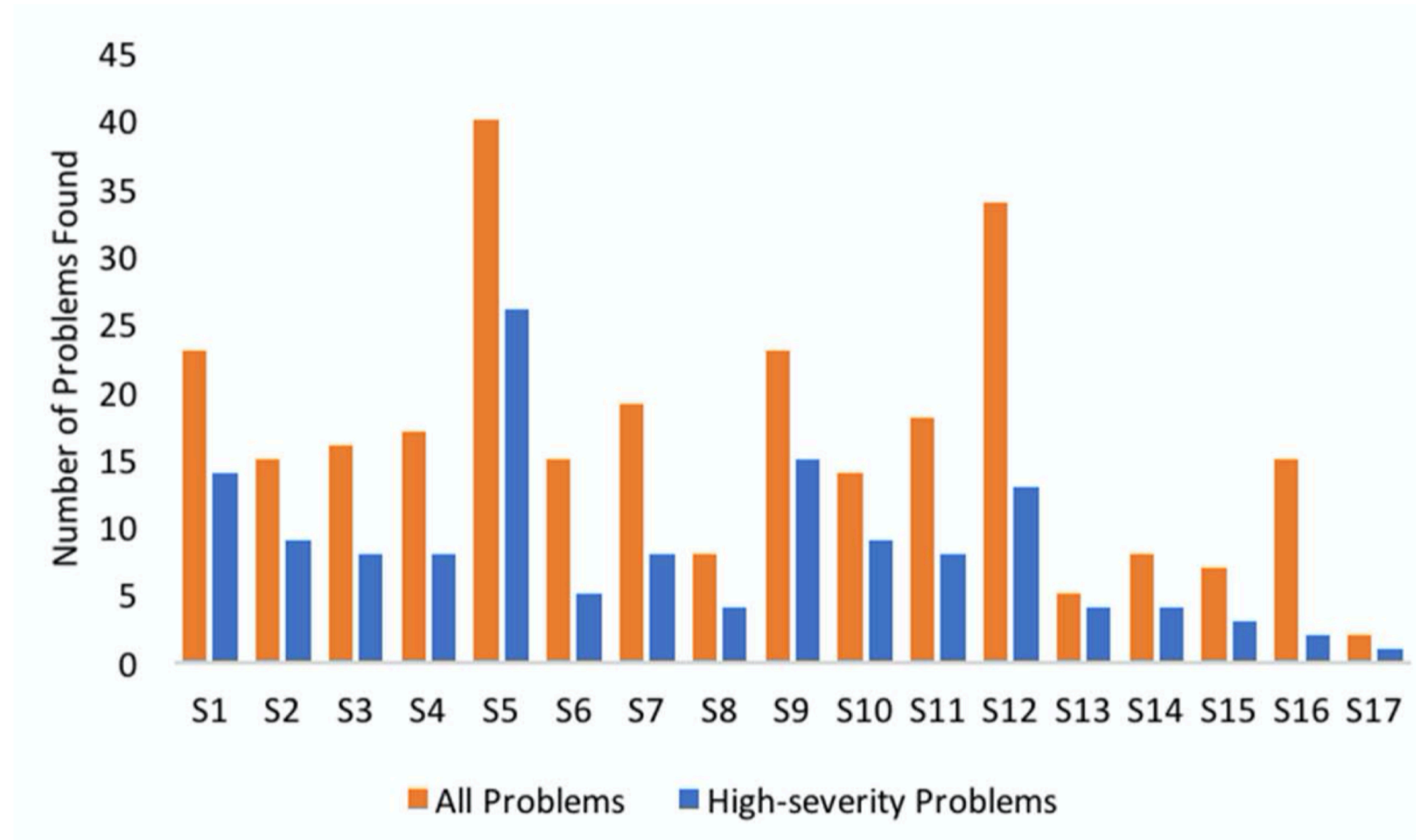
### S16: Use normal language in communicating errors.

- Vary (error) prompt wording on re-prompts.
- Don't blame the user for errors (don't say: "that was not a valid response").
- Don't show mock concern (don't say: "I'm sorry. I did not understand the response I heard.").

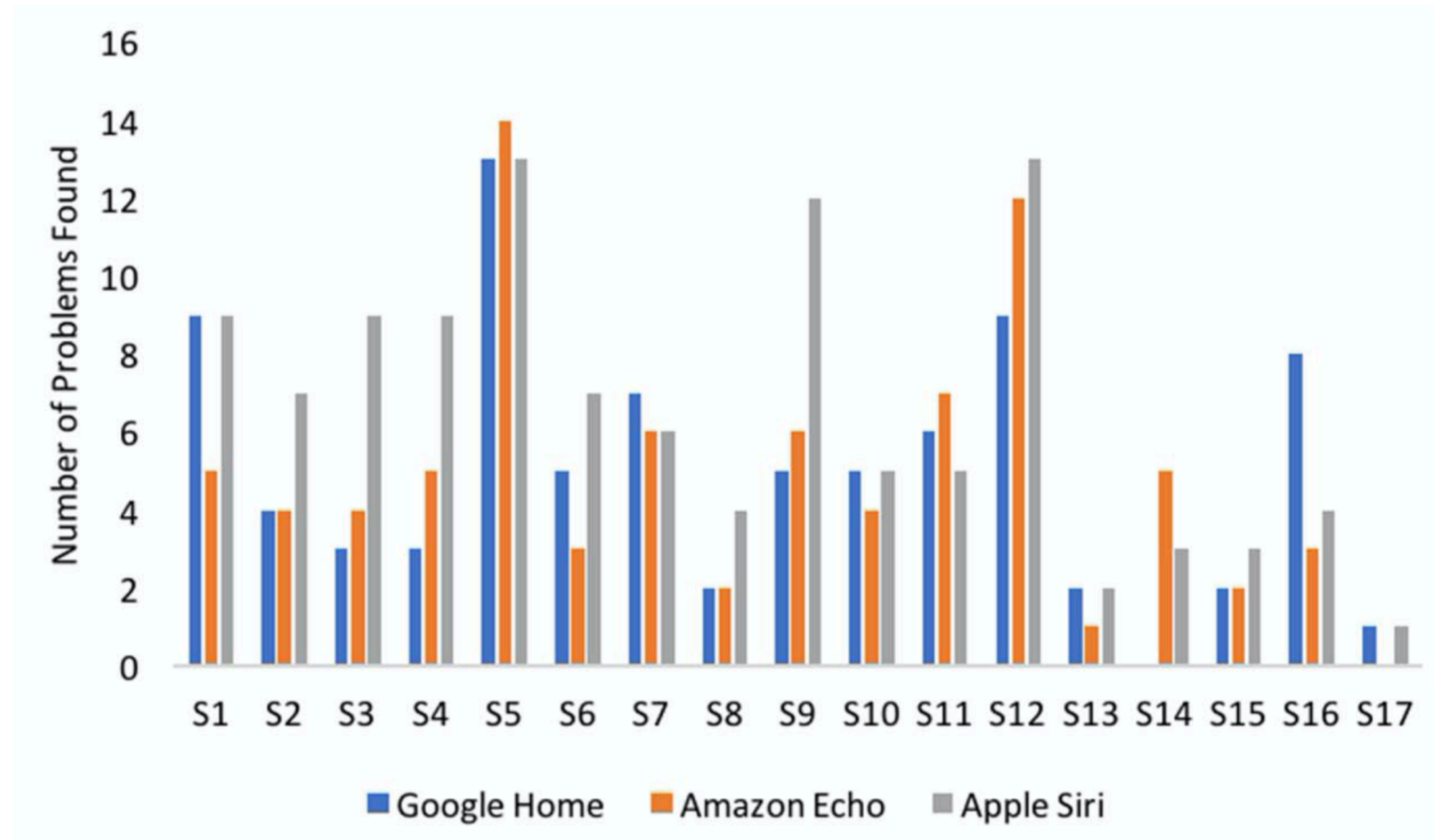
## Heuristic #17

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

- Use a special escape word globally (e.g., “Stop”).
- Use non-speech methods when speech fails (e.g., push a physical button).



<sup>7</sup>Wei & Landay, 2018, Speech-based Conversational Agent Heuristics



<sup>7</sup>Wei & Landay, 2018, Speech-based Conversational Agent Heuristics

# What did we learn today?

- Design principles for conversational interfaces
- Usability heuristics for conversational interfaces