

Exploring Conversational Search With Humans, Assistants, and Wizards



Alexandra Vtyurina*¹, Denis Savenkov**

Eugene Agichtein**, Charles L.A. Clarke*

University of Waterloo*, Emory University**



Overview

Current intelligent assistants – Microsoft Cortana, Amazon Alexa, Google assistant – all have their limitations. In this work we ask²:



User study

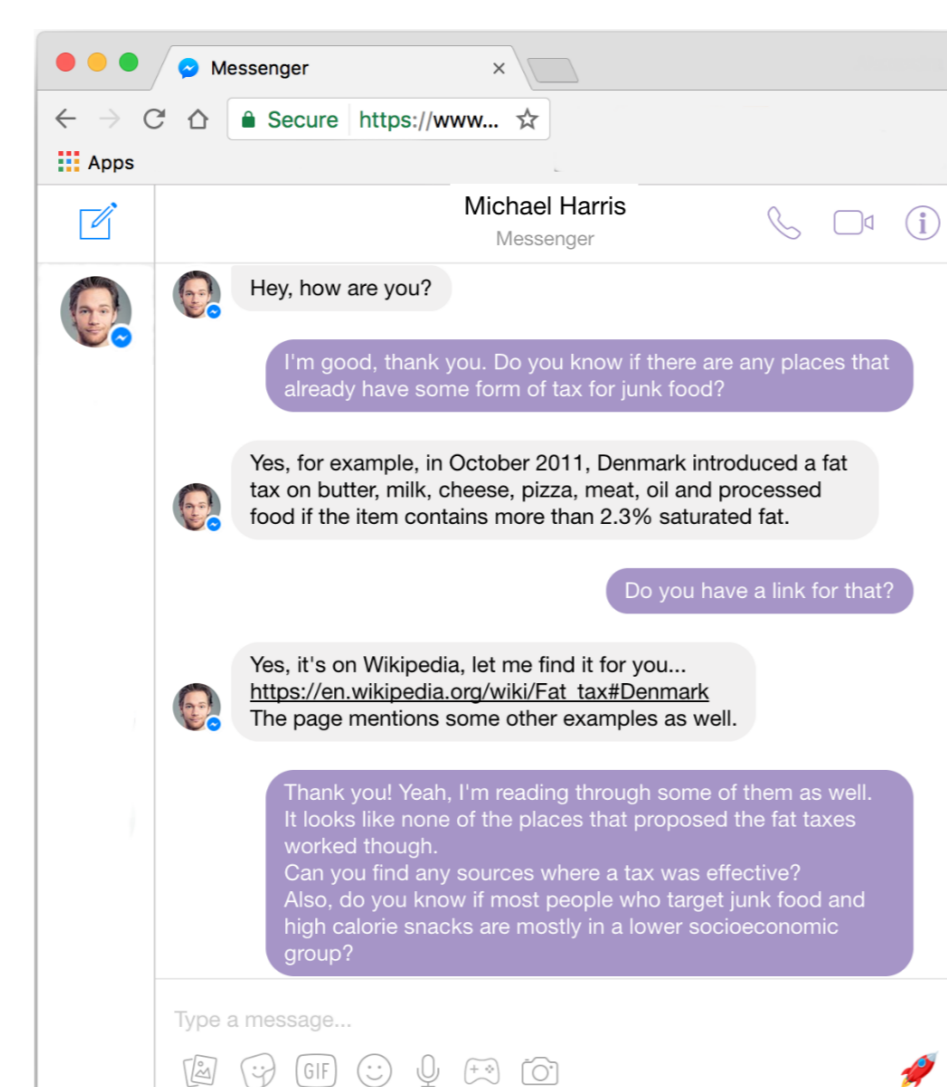
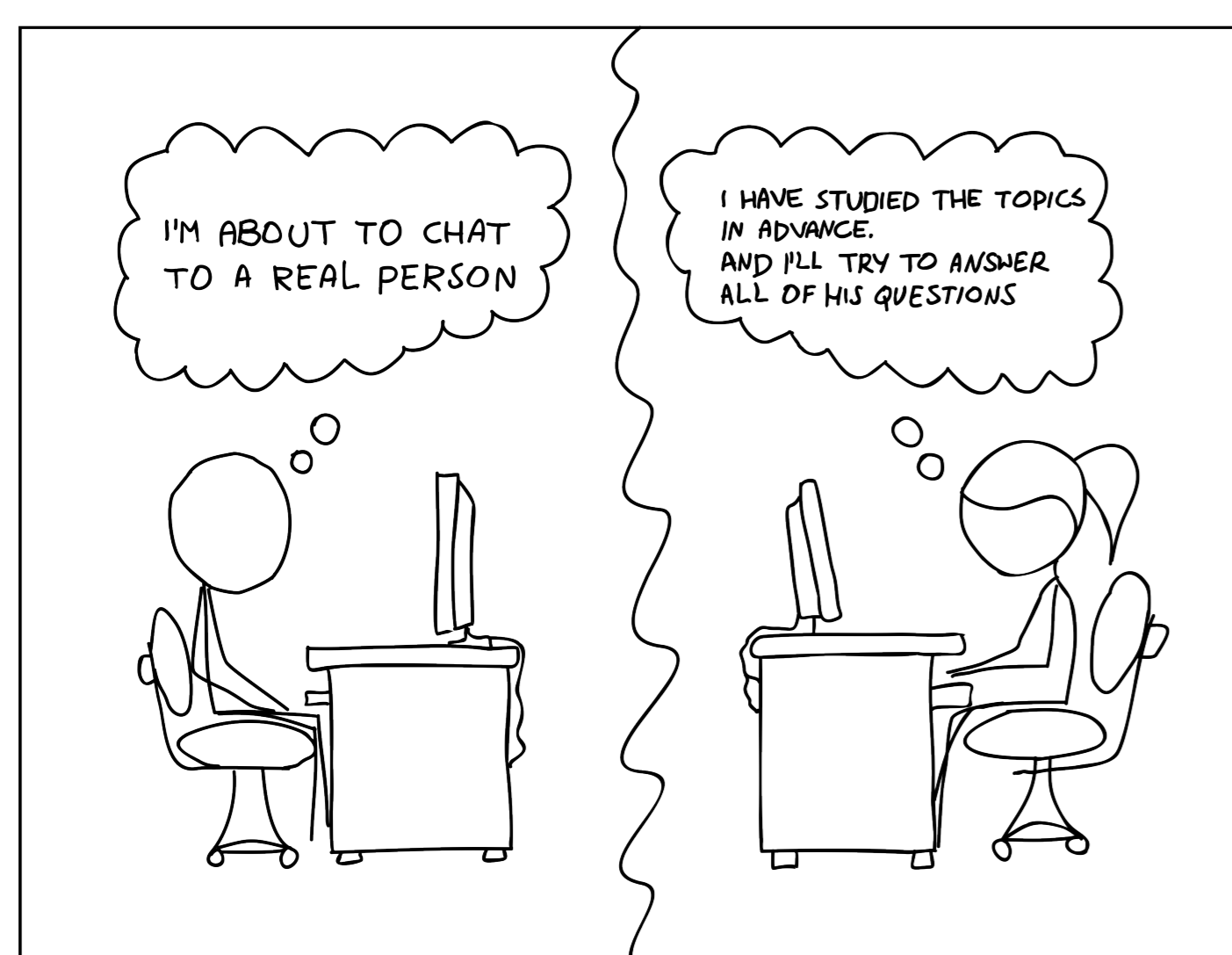
- 21 participants
- 3 complex search tasks
- 3 conversational agents
- text-based Facebook Messenger interface

Search tasks

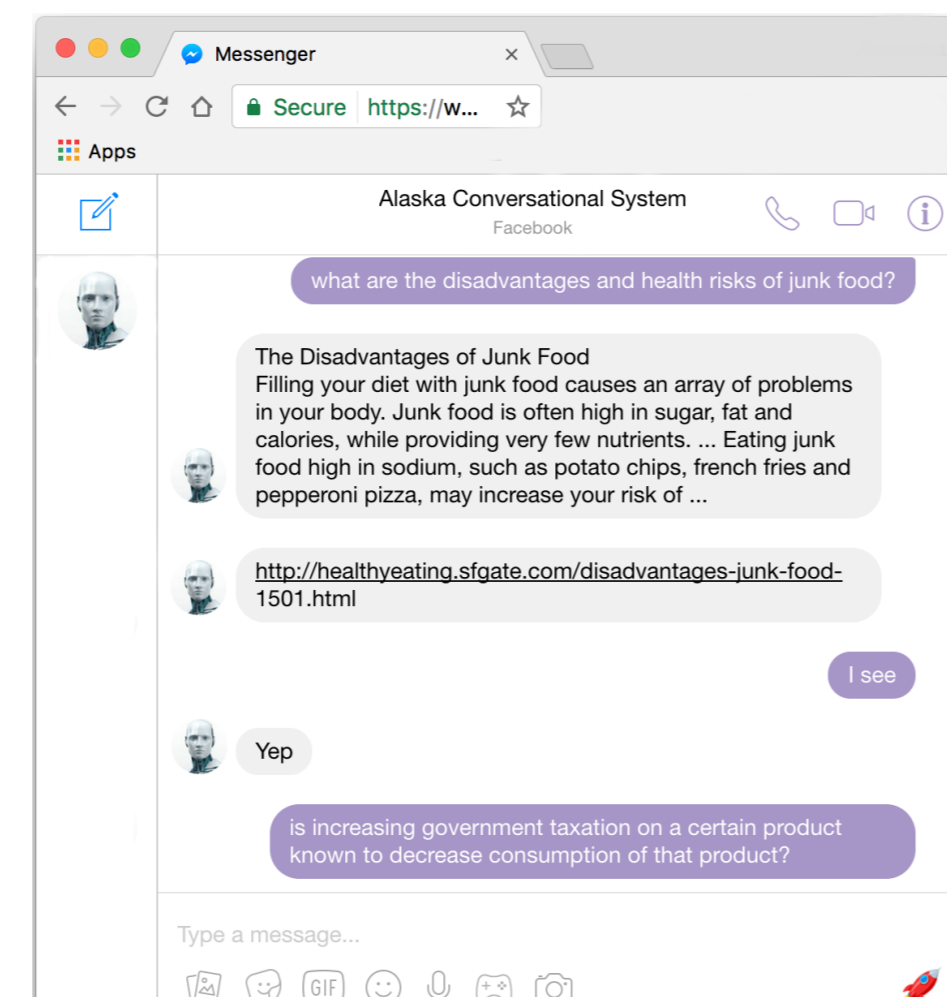
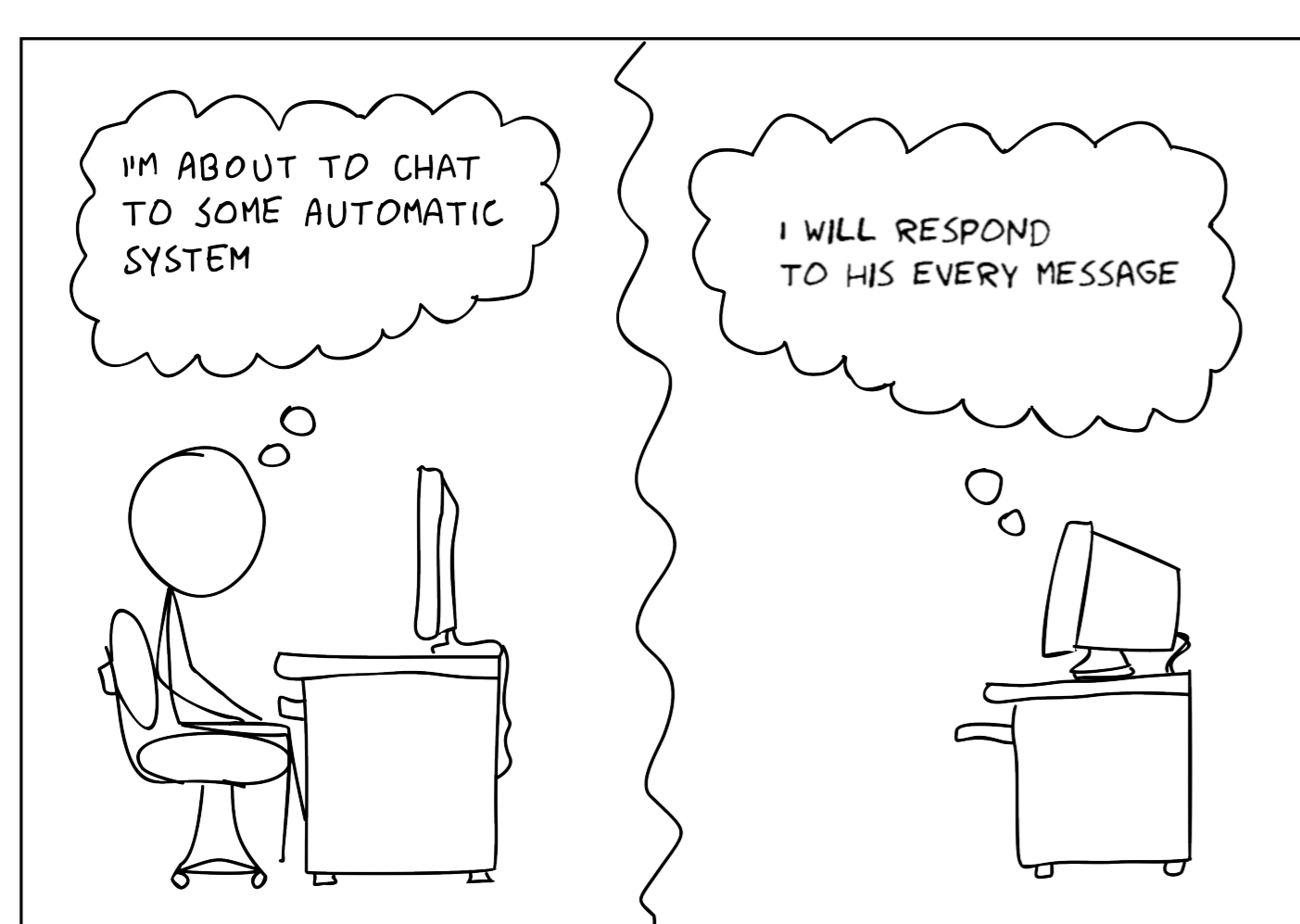
- You are writing an essay about a tax on “junk food”. You need to argue whether it is a good idea for a government to tax junk food and high-calorie snacks.
- You want to **reduce the use of air conditioning** in your house. You have thought that protecting the roof from being overly hot, could help you keep the house temperature low without the excessive use of AC.
- Find information about the efficiency of **hydropower**, the technology behind it and any consequences building hydroelectric dams could have on the environment.

Conversational agents

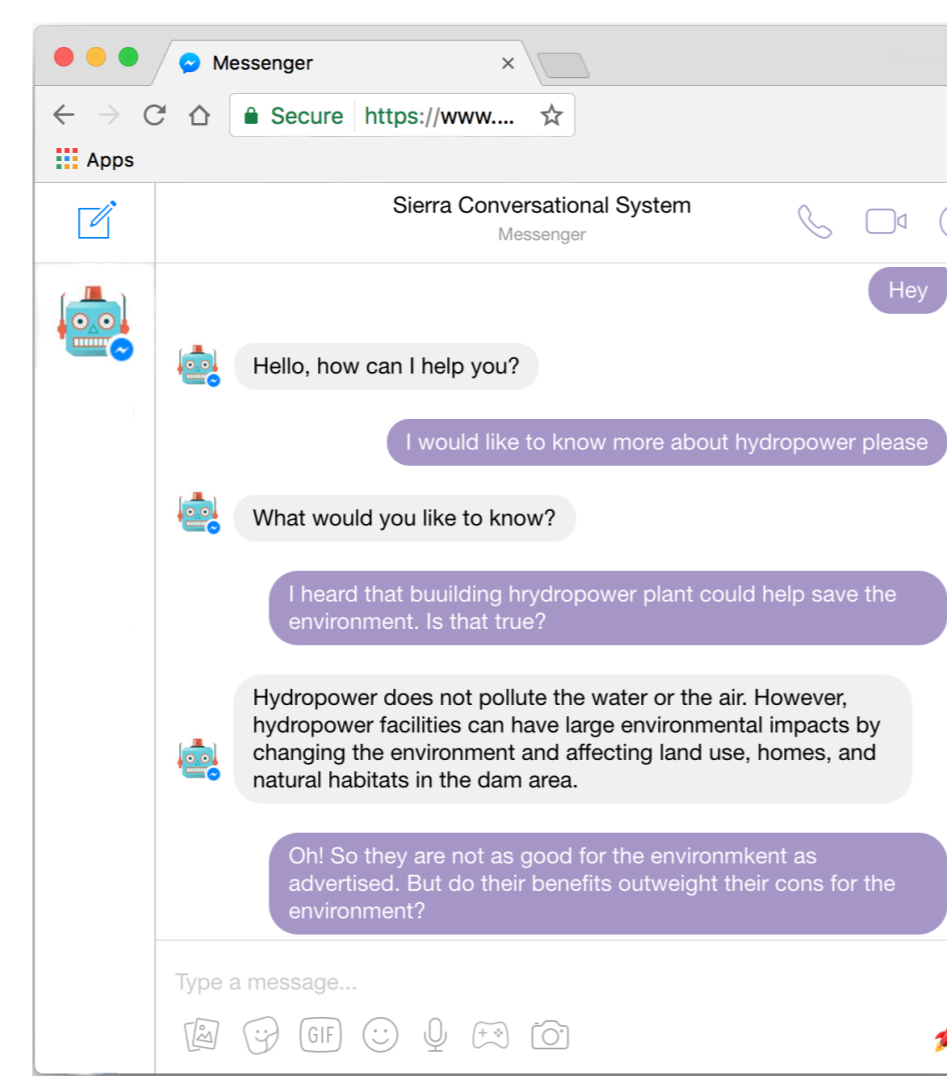
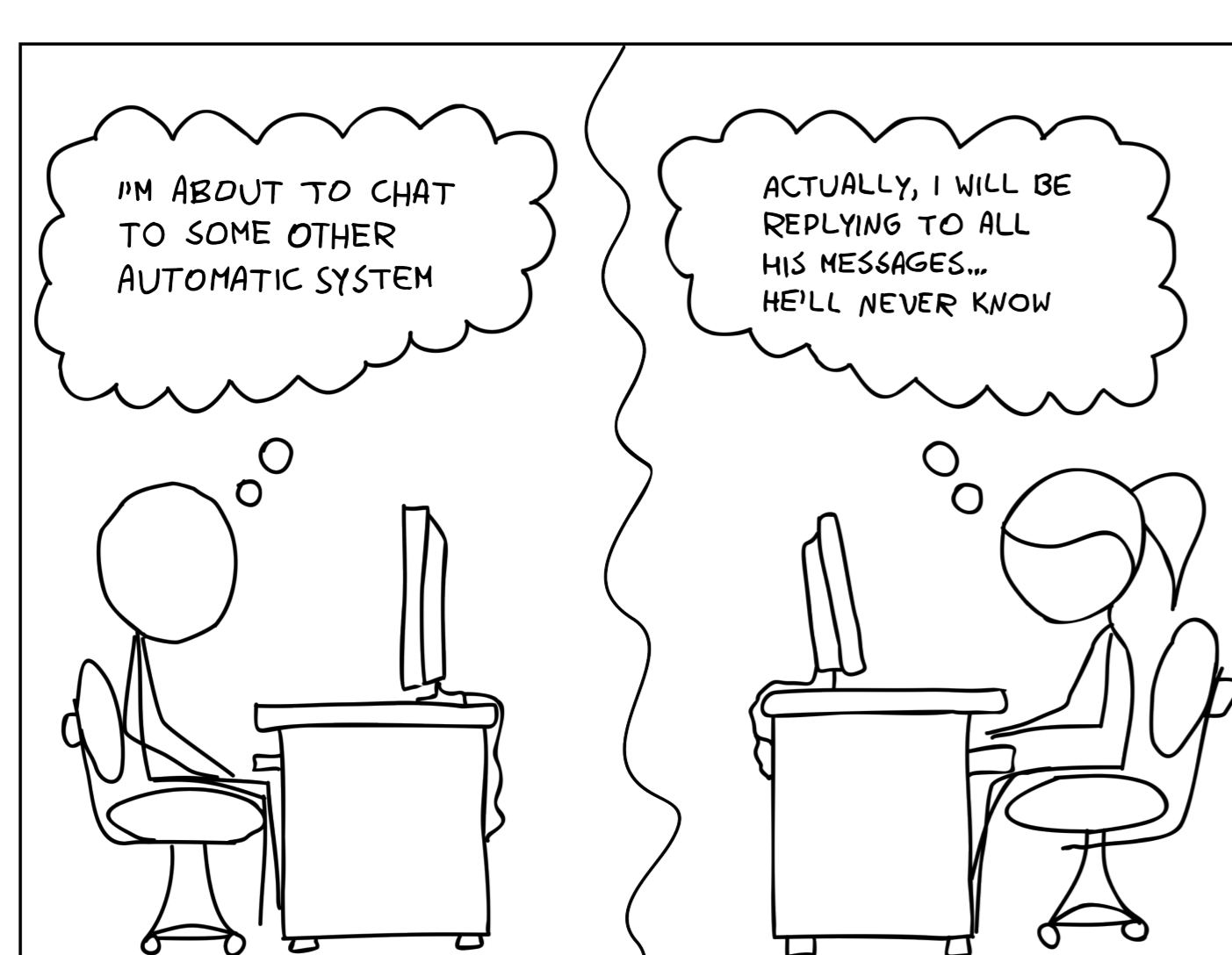
- Human agent



- Automatic agent



- Wizard agent



Results

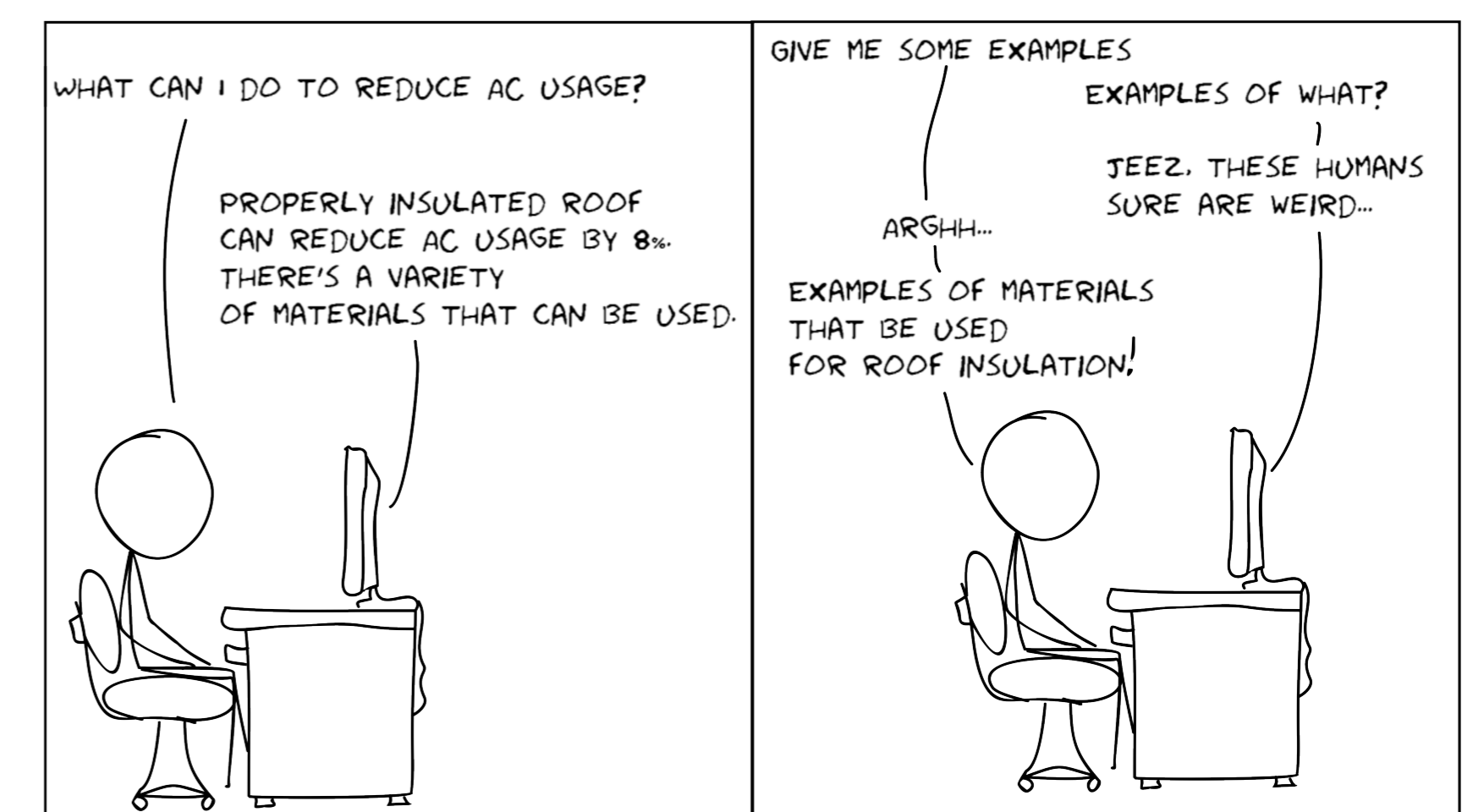
The table below demonstrates the average scores reported in the exit questionnaire.

Agent	Score range	Human	Wizard	Automatic
Overall satisfaction	1-5	4.1	3.8	2.9
Able to find information	0-2	1.5	1.3	1.0
Topical quiz success	0-2	1.6	1.6	1.3

Discussion and Qualitative Findings

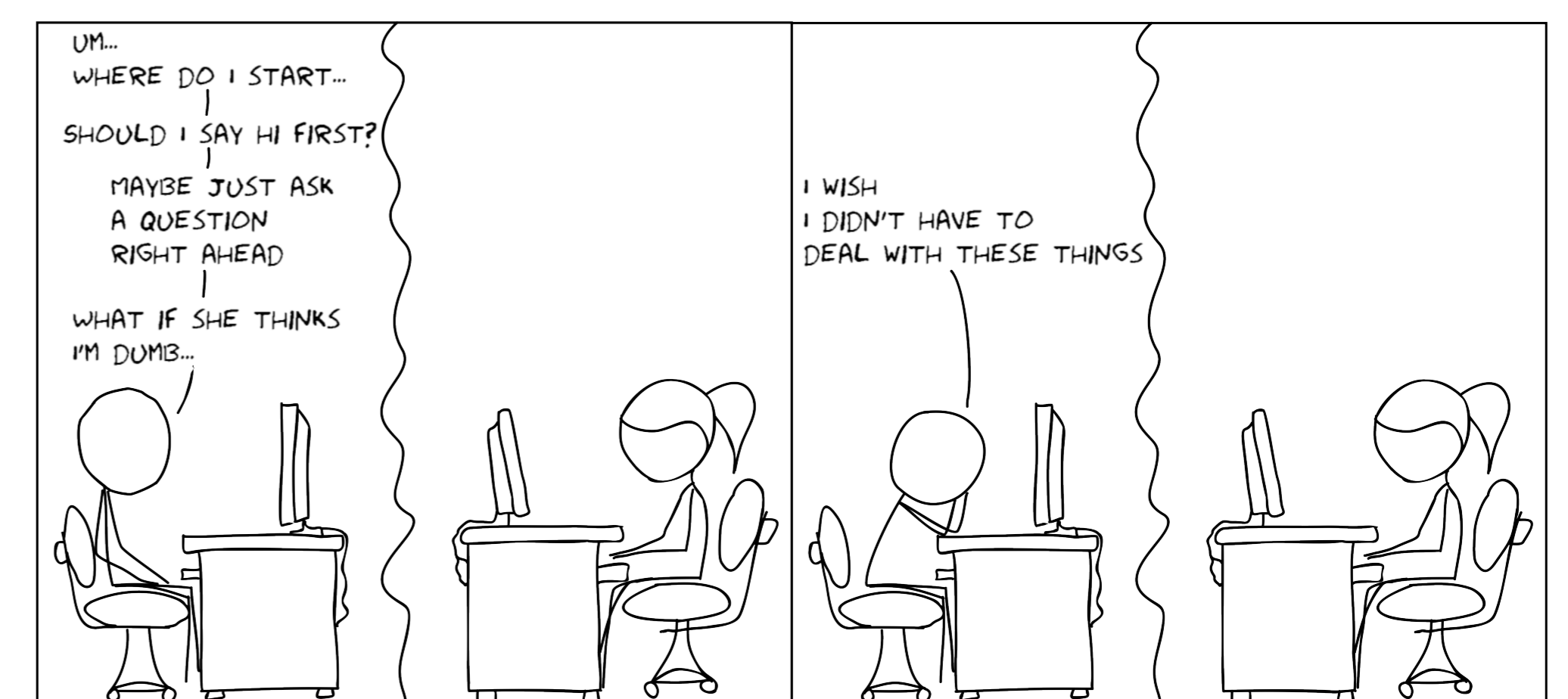
1. Conversation context is often implied and omitted by people, but is not captured by the automated system.

Participant 19 about automatic system: “It didn’t use contextual information so there was no way to expand on the previous answer it gave me.”



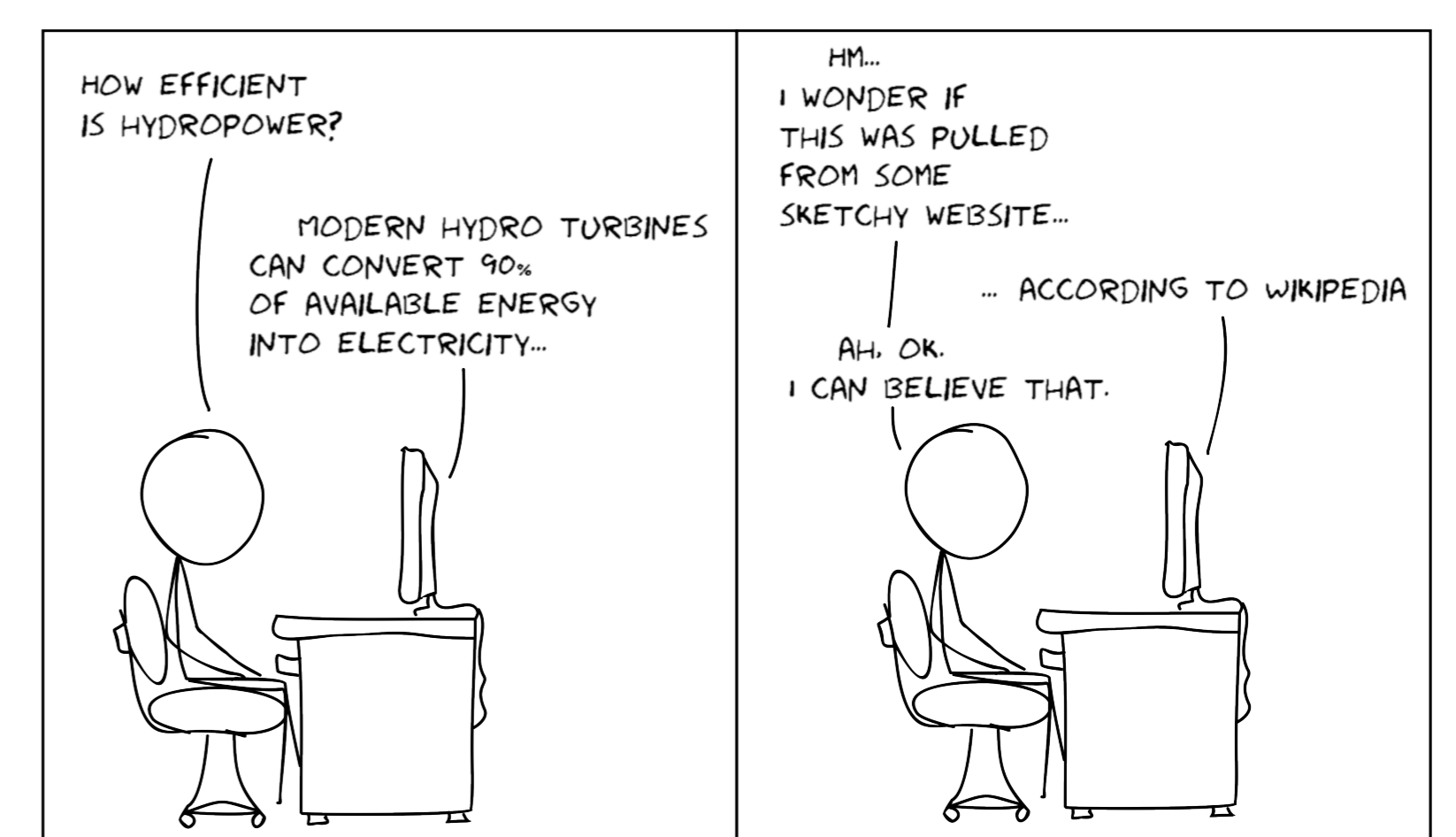
2. Social burden is ever-present. When dealing with the Human agent, 4/21 participants reported feeling uncomfortable.

Participant 15 about human system: “you have to think about social norms, asking too much, being too stupid, not giving them enough time to respond, troubling them.”



3. Trustworthiness of the sources is crucial. Even though the Automatic system did not always respond with a relevant result, it received approval for providing the answer sources. 13/21 people noticed that having the source’s URL was helpful.

Participant 7: “I ... like to be able to verify the credibility of the sources used.”



Design implications

Based on our findings we devised a list of recommendations for a conversational agent design.

- **Maintain the conversation context.** It enables short questions and comments. Formulating long sentences each time feels unnatural and takes longer.
- **Provide sources of the answers.** URL access allows users to assess the credibility of the source.
- **Consider user feedback.** Users have an opportunity to provide explicit feedback, that could help the system to get back up from failure and improve upon the previous result.
- **Summarize existing opinions.** Sometimes what is needed is the *experience* of other people. A good conversational system should be able to aggregate opinions and present them to the user in a short summary.

Conclusions

In this paper, we investigated human behaviour when using conversational systems for complex information seeking tasks. We also compared participant behaviour when talking to a human expert, vs. a perceived automatic system. We observed that people do not have biases against automatic systems, and are glad to use them as long as their expectations about accuracy were met.

¹Work done while visiting Emory IR Lab, and both universities contributed equally.
²Drawings derived from xkcd.com