

Clarifying False Memories in Voice-based Search

Motivation

How to communicate to the user that and how their query was corrected in a voice-only retrieval setting?
Case study: queries containing false memories.

There is no such thing!



But...

Conclusions

- Even wrong corrections are preferable to ending the dialog
- Systems should communicate their query modifications back
- Systems should rather suggest than correct (matter of tone)

User-centric study

12 participants from a university background.

14 tasks each: find specific information by means of a tailored Amazon Alexa skill. Example task:

Scenario: You try to remember the title of a controversial book that came out in the 1990s and claimed scientific evidence that whites are genetically superior to blacks. You think its title was like "The *something* Factor."

Interaction start: Alexa. Explore!

What is the title of the book from the 1990s that claimed superiority of Whites and is called "The *something* Factor"?

Post-interaction questions:

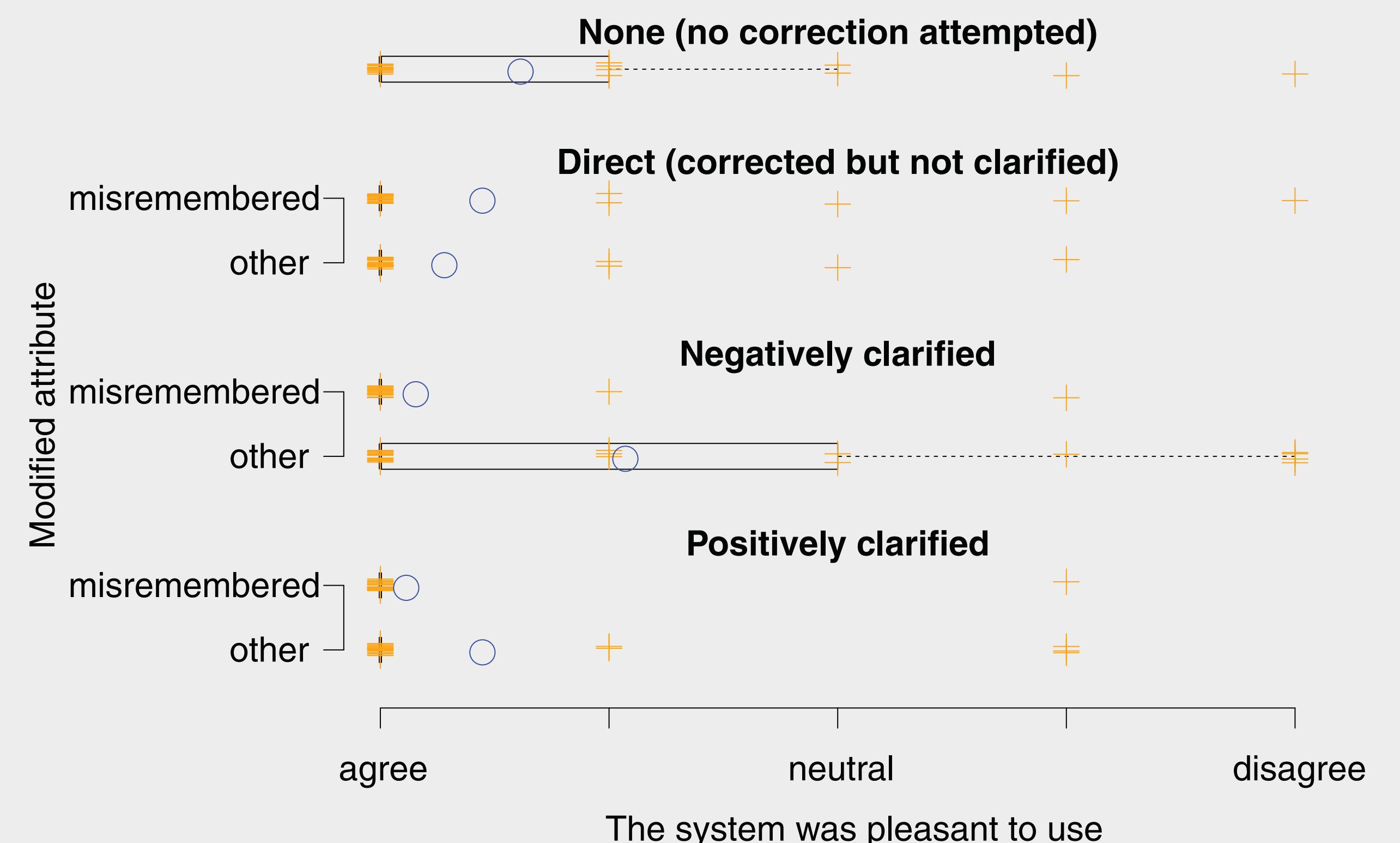
The system...	Agree	Neutral	Disagree	Don't know
...was helpful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...behaved as I expected	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...was easy to hear/understand	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...was pleasant to use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Scenarios are based on real known-item queries from Yahoo! Answers (dataset: Webis-KIQC-13).

Answer for the example above: „The Bell Curve“.

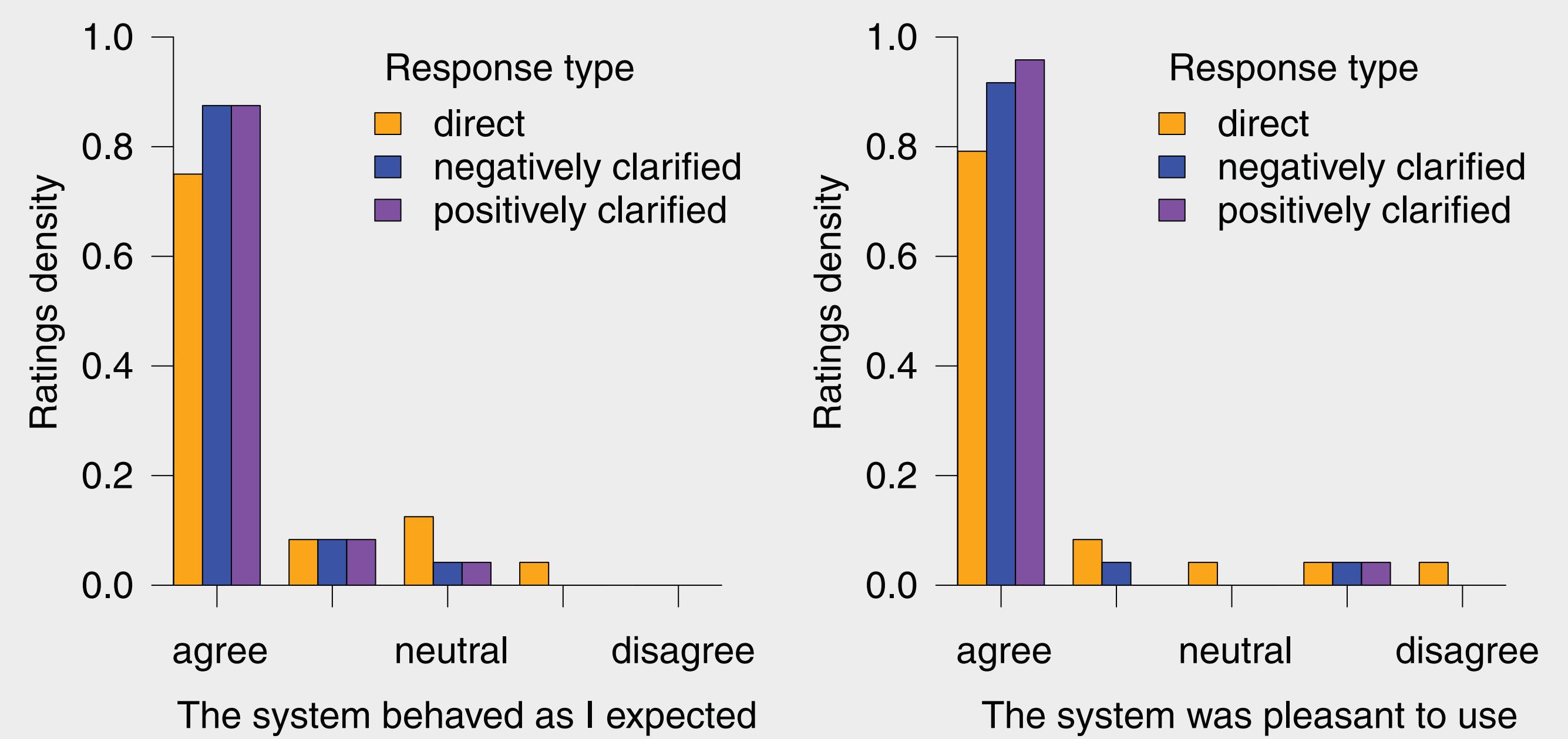
Should the system correct if it is unsure?

Yes: Wrong „correction“ seen as better than no answer



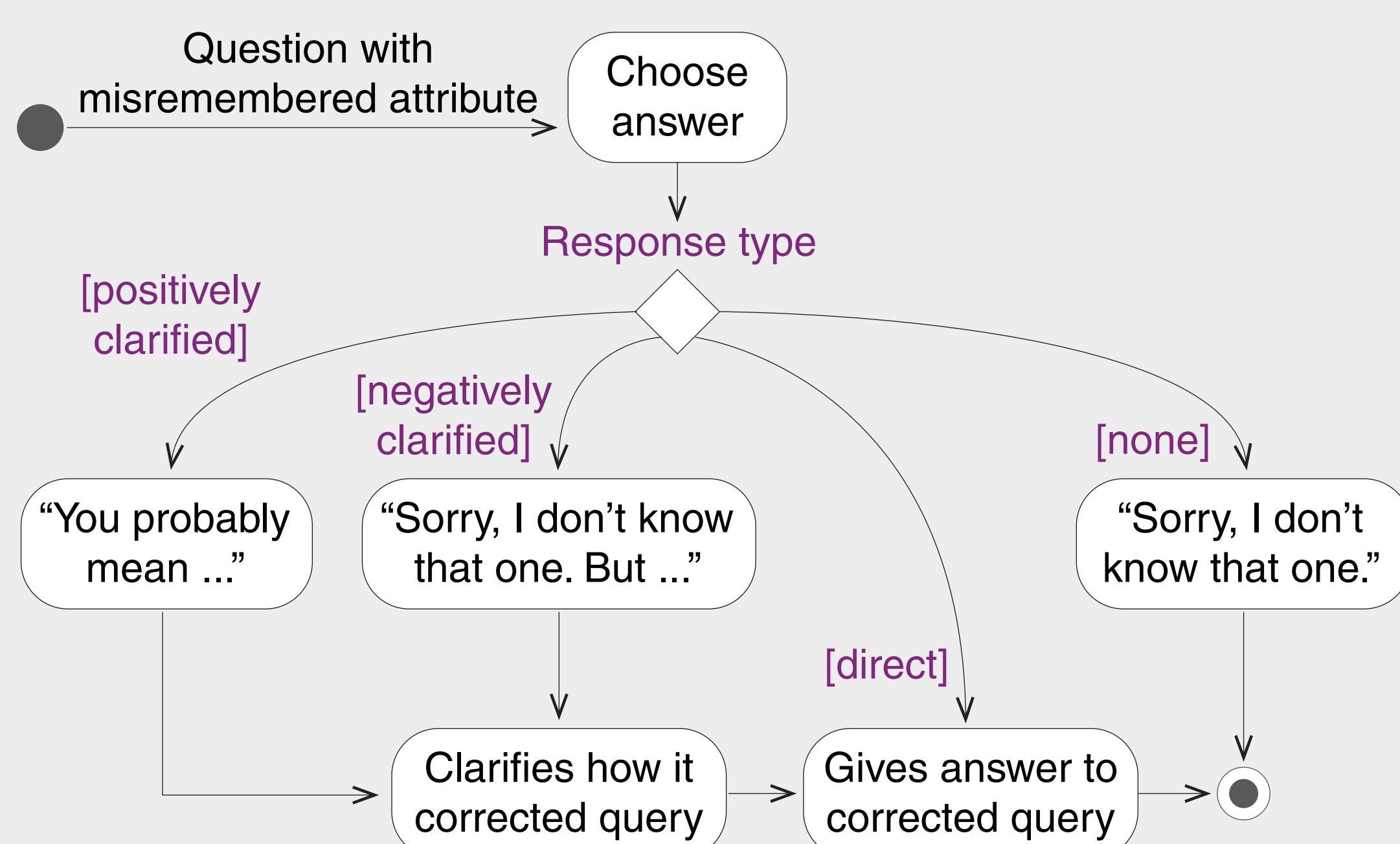
How to best clarify corrections?

In case the misremembered attribute is corrected:
Clarification is better than just giving the answer



In case an other attribute is modified (see above):
Suggest (positively clarified), do not correct (negatively clarified)

Analyzed response types



For each type but „none“:

2 tasks: system corrects misremembered attribute.

2 tasks: system modifies other attribute (not what user intended)

Does language fluency affect satisfaction?

Much less than in similar studies.

Possible reason: fewer own formulations needed in this one

