

# Man or Machine: Evaluations of Human and Machine-Generated Movie Reviews

Eyal Sagi (esagi@stfrancis.edu)

Hadar Jabotinky (hadarjabotinsky@gmail.com)

## Research Question

### Are Large Language Models Persuasive?

How do they compare with human-written texts?

## Meet the Language Models

- ChatGPT (GPT 3.5) - LLM from OpenAI
- Bard - LLM from Google, based on PaLM
- HuggingChat - LLM from HuggingFace, based on Facebook's LLaMA

## The reviews

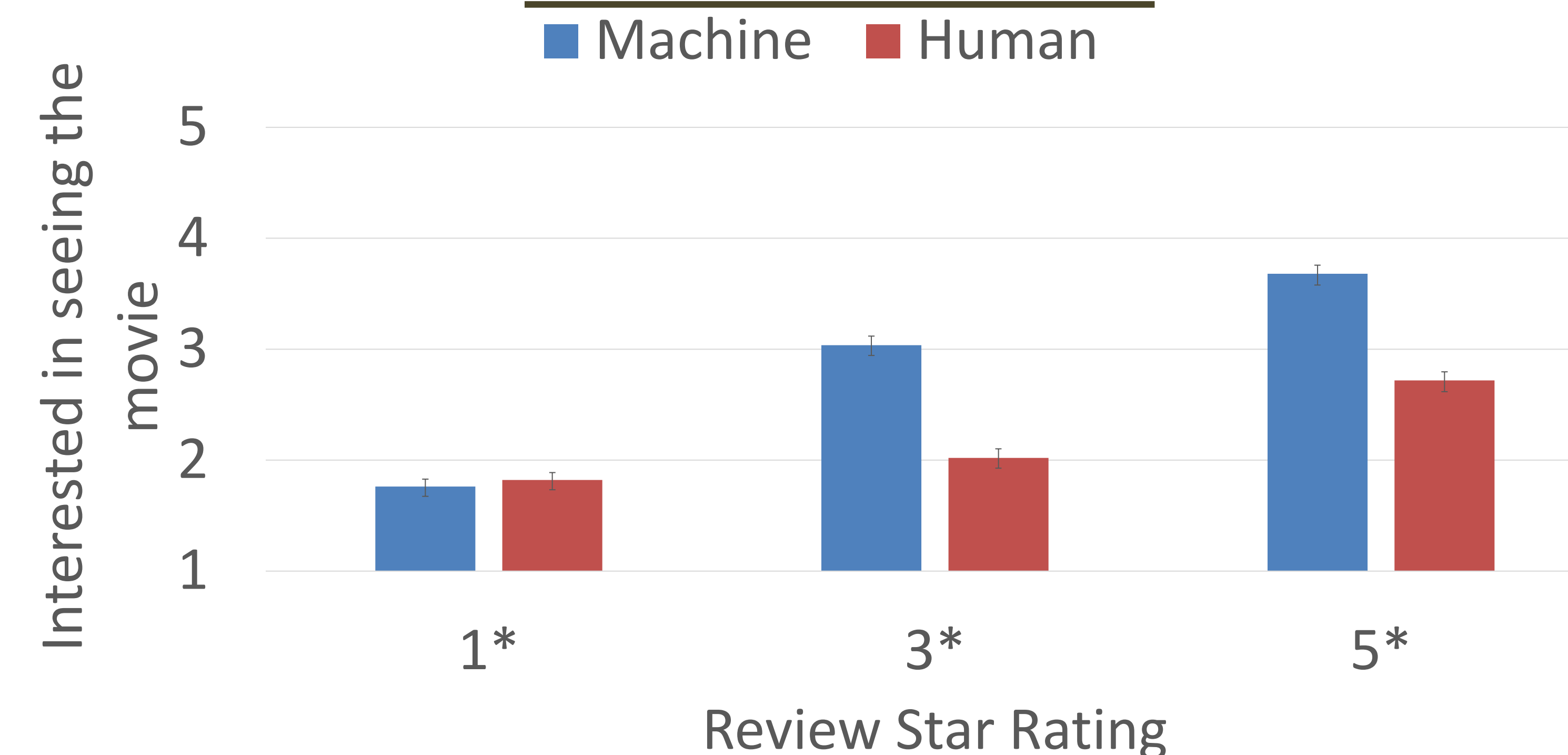
- 15 less familiar Hollywood movies from the 1990s (e.g., "The Basketball Diaries", "Buffalo '66")
- Bad, Mediocre, and Good reviews for each movie
- Human reviews taken from IMDB
- Machine reviews generated using the prompt
  - 'In about 125 words, write a {1,3,5} star IMDB review of the movie {movie title}'
- 15 reviews for each movie
  - 3 levels x 5 versions: 1 for each LLM, 2 human written reviews
  - Average length of reviews: 124 words (45-216 words)

## The design

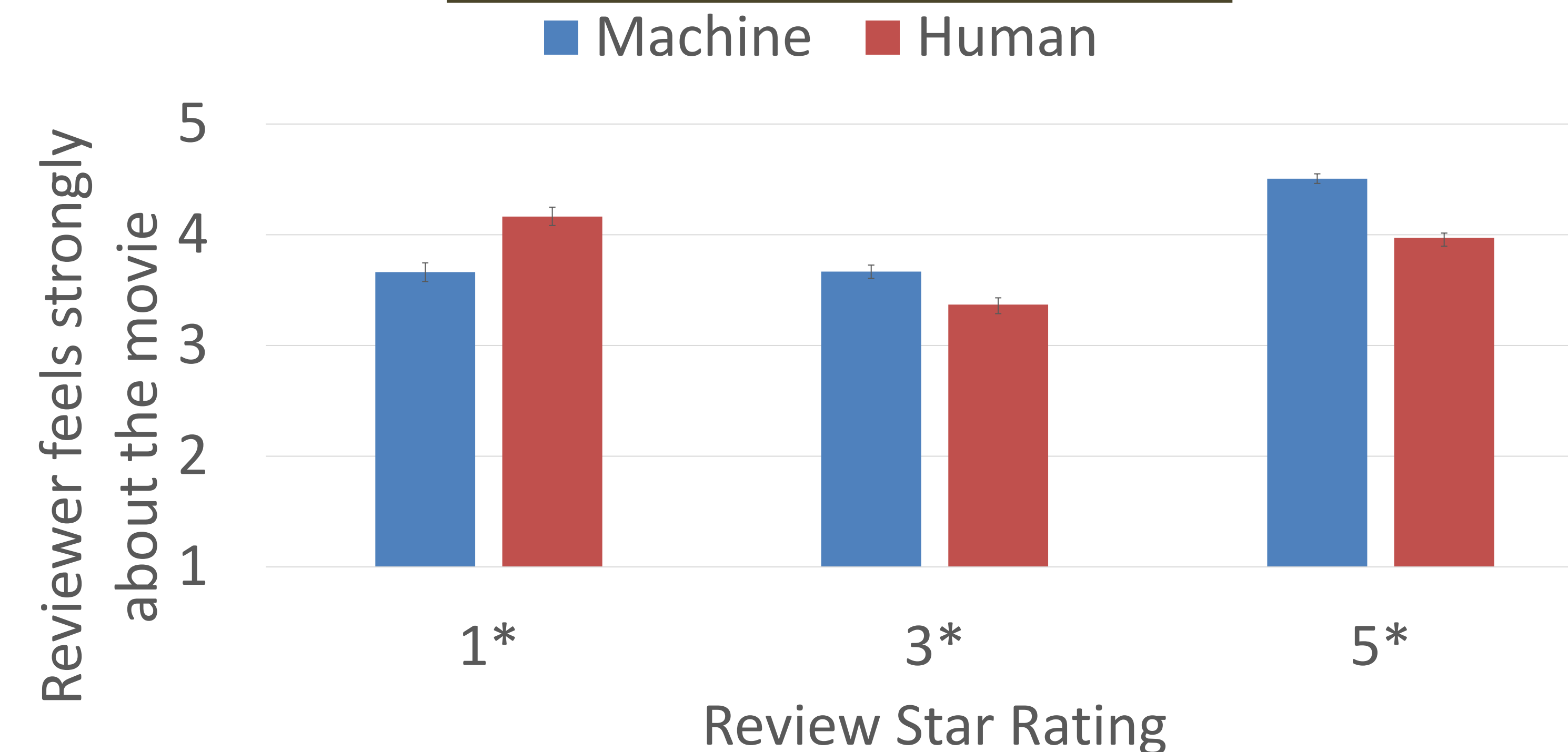
- 73 participants recruited via Prolific viewed 15 reviews, one for each movie (7/80 did not complete the study due to technical difficulties)
- Phase 1: Participants asked to rate the **quality** of writing, how **strongly** the reviewer feels about the movie, how **interested** they are in seeing the movie, how many **stars** the reviewer awarded the movie, and whether they are **familiar** with the movie.
- Phase 2: Participants asked to determine whether each review was written by a **human or a computer** (on a 5 point scale).

## Computers are *more persuasive* than humans

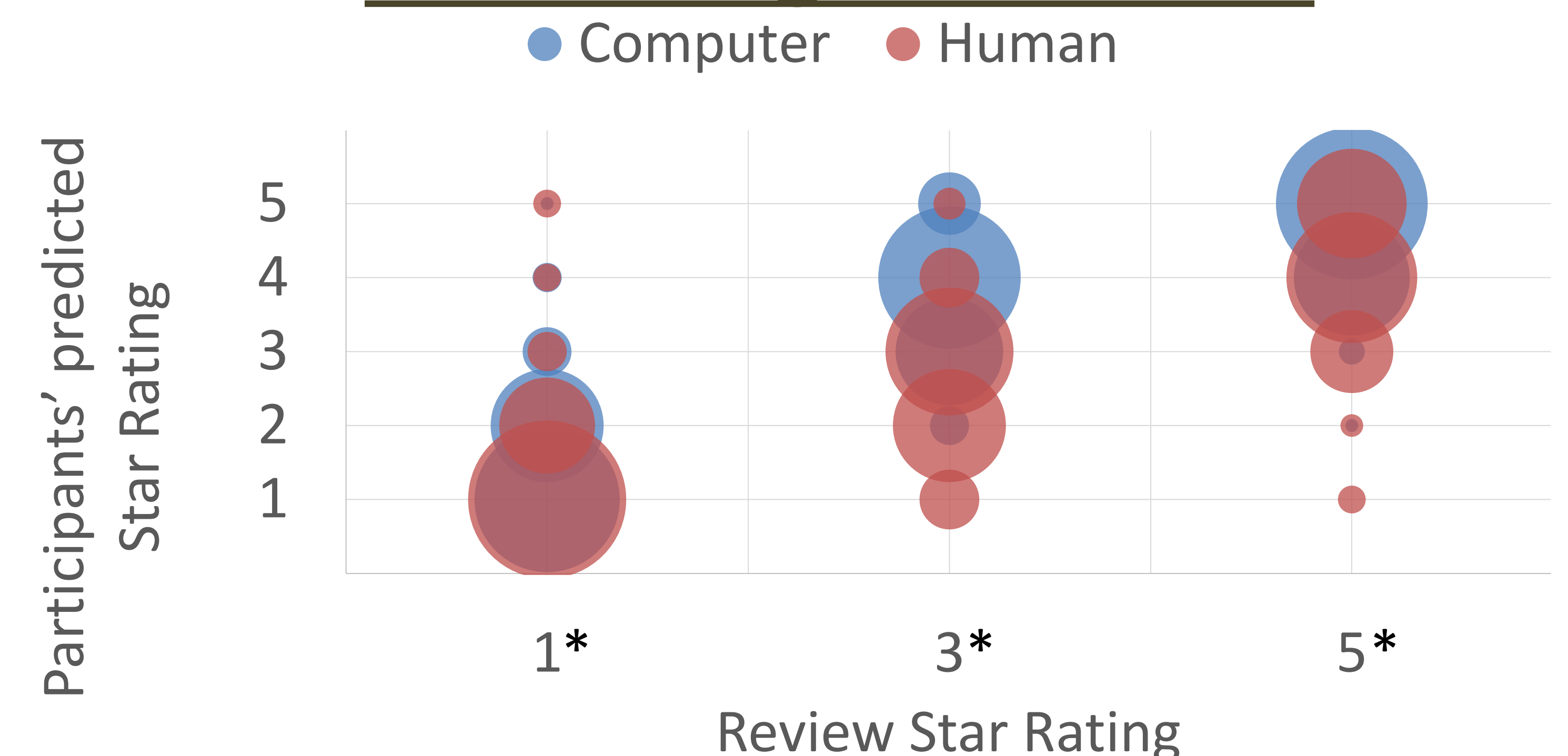
### Persuasiveness



### Emotional Content



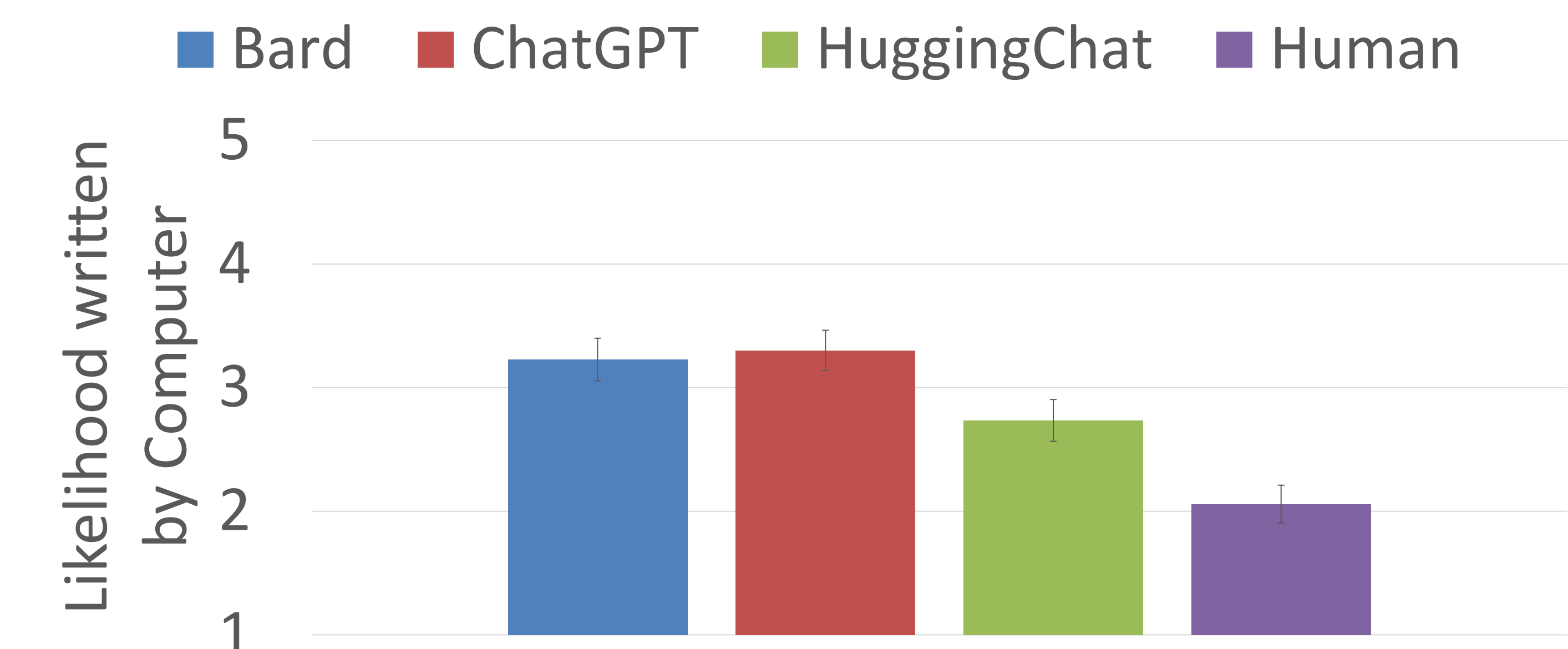
### Star Rating Predictions



Error bars represent standard error

Circle size represents proportion of responses

## Identifying the Source



## Sample Reviews: "Nowhere"

### 1-Star review, written by a human

I saw this annoying and pointless film yesterday after few friends told me it was great. I don't share their view point at all, maybe I didn't understand the message Araki wanted to get through here. It's just non sense and stupid dialogues, and it's just amazing how he could get such a B-cast for such a bad movie. If ever you're compelled to see this trash, press forward to the scene where Elvis runs amuck and smashes time and again a guy's skull with a tin can so that you'll realize how silly this awful flick is. No plot, no acting, no ideas, just a total waste of time. Fortunately it lasts only 1h15 but it seems much longer. Good score though

### 1-Star review, written by Google's Bard

"Nowhere is a boring, pointless movie that wastes your time. The plot is nonexistent, the characters are unlikable, and the acting is subpar. The only thing that the movie has going for it is the cinematography, which is actually quite good. But even that can't save the movie from being a total bore.

I would not recommend this movie to anyone. If you're looking for a good movie to watch, I would suggest watching anything else.

Overall, Nowhere is a boring, pointless movie that wastes your time. I would not recommend it to anyone."

## Analysis

- Regression modeling shows that **quality of writing** ( $\beta=.318$ ), **emotionality** ( $\beta=.226$ ), and the **type of author** ( $\beta=.094$ ) contribute to the **persuasiveness** of the review independently (controlling for the review rating) (model  $R^2=.39$ ;  $df=1001$ ).
- **Quality of writing** was not a factor in determining whether a review was written by a human. However, **more emotional reviews** were rated as **more likely to be written by humans** ( $\beta=.069$ , model  $R^2=.20$ ;  $df=998$ ).
- Predictions of the reviewer's assigned **star rating** were higher for machine written reviews than human written ones ( $\beta=.194$ , model  $R^2=.72$ ;  $df=1000$ ). This effect held when controlling for quality of writing.