



Scan to see digital poster

Finding the Conceptual Glue: Ad-hoc Categorization in People and LLMs

Aja Altenhof*, Kira Breeden*, Gary Lupyan



Categorization

Conventional

A stored representation with readily available explanatory framework

BIRDS

DUCK, OWL, RAVEN, GOOSE

Ad-Hoc Schema-Based

Ad-Hoc categories require people to link seemingly “unrelated” concepts (Barsalou, 1983)

THINGS TO SAVE IN A FIRE

BIRTH CERTIFICATE, PET, FAMILY PHOTO, A RING

“Truly” Ad-Hoc

Learners must find the “conceptual glue”

THINGS AESTHETICALLY MOTIVATED GIRLS POST TO INSTAGRAM

STREET, COFFEE, FOOD, HAIR

Materials

90 word sets

M semantic similarity = 0.34

M concreteness = 4.20

M frequency = 76.098

STREET

COFFEE

FOOD

HAIR

Words were randomly selected from Wordnet (Miller, 1995) and matched for frequency and concreteness within groups

Procedure

Category Generation

Deepseek-chat
GPT 5.1 Thinking

N = 55

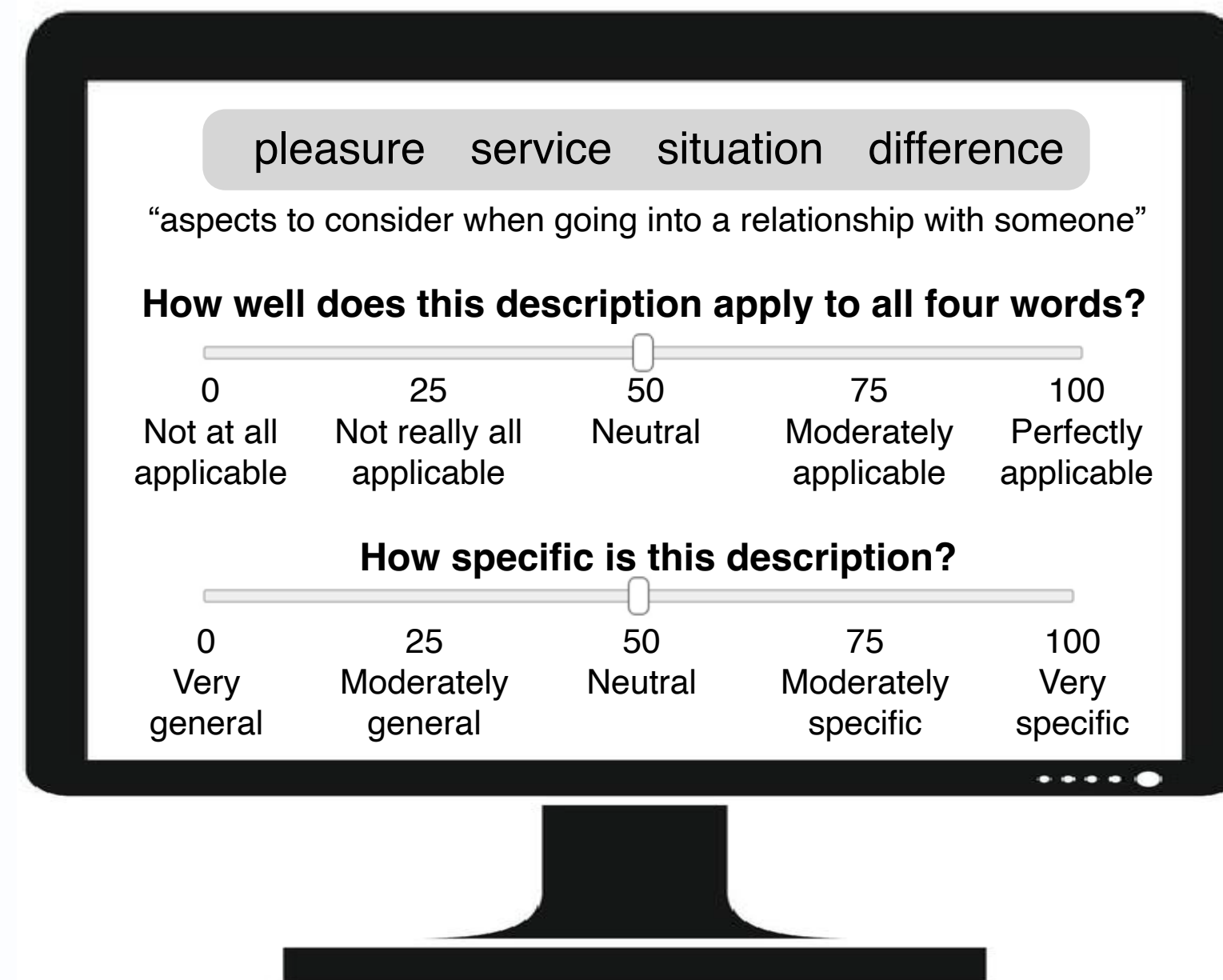


How difficult was it to come up with this category?
What percent of people do you think would generate the same category you did?

1962 categories
~18-22 per word set

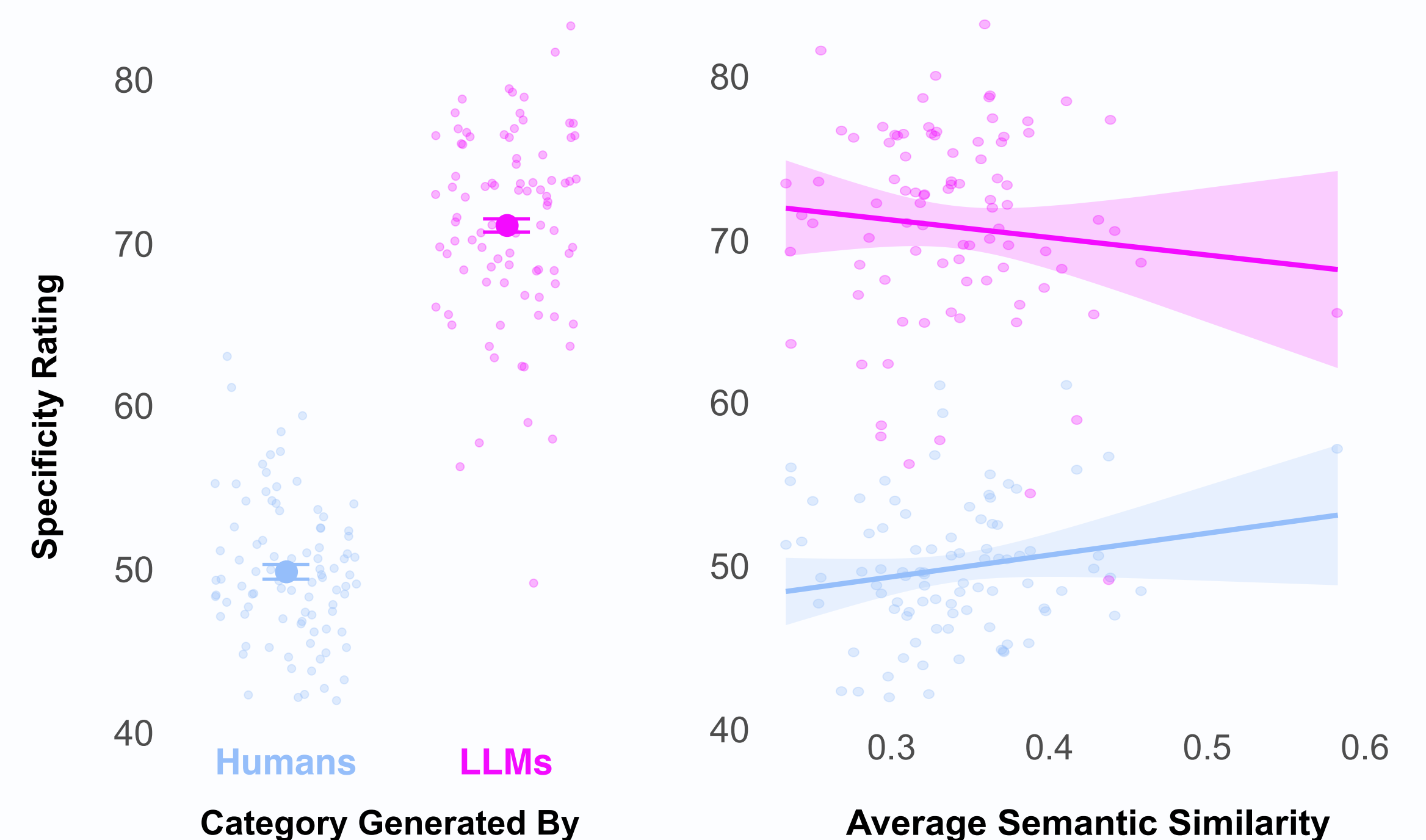
Category Applicability & Specificity Ratings

N = 415



M = 10 applicability and specificity ratings per word set

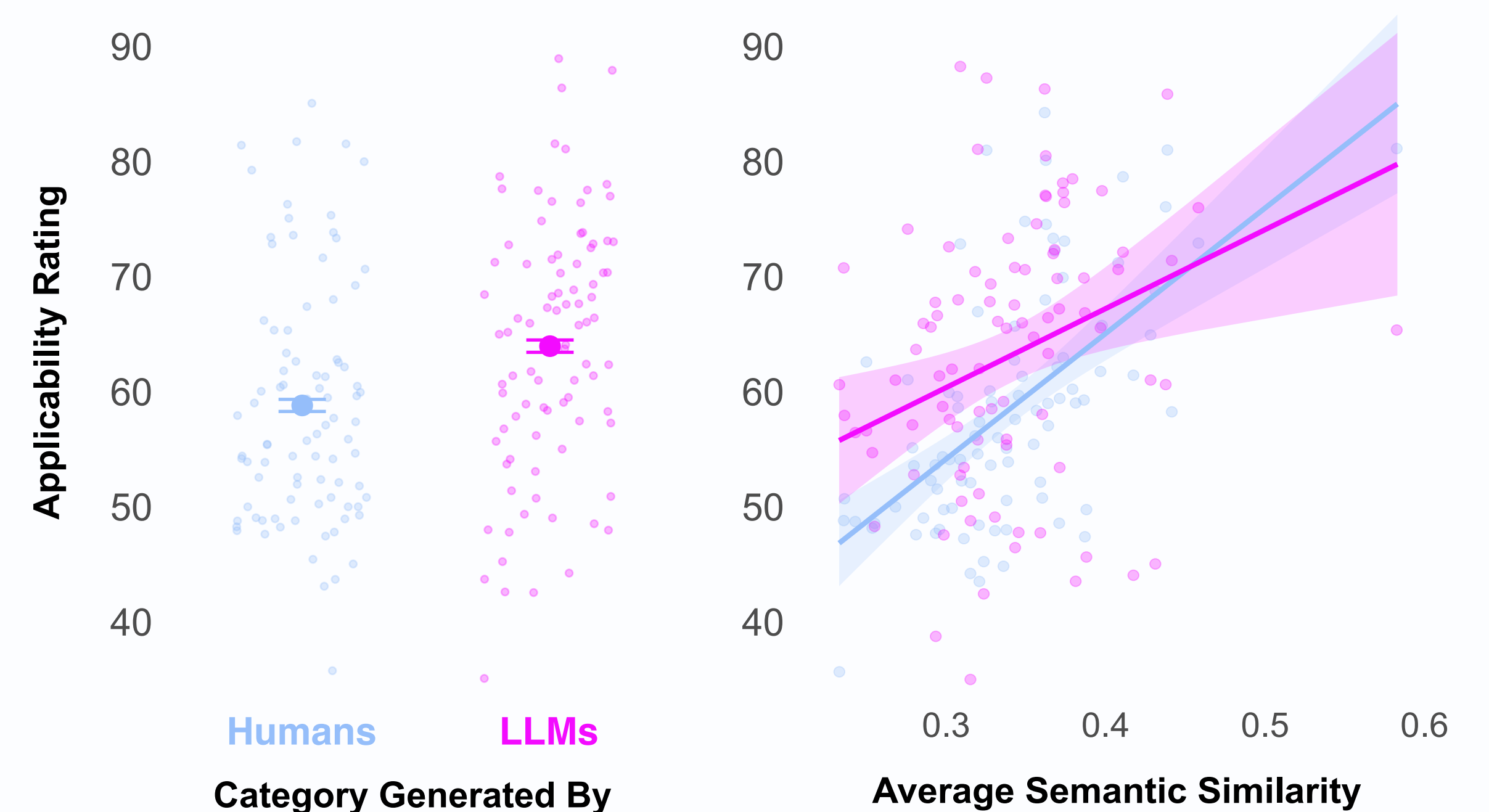
Category Specificity



LLMs generate more specific categories than people

People have an especially hard time generating specific categories when sets are composed of less similar items.

Category Applicability

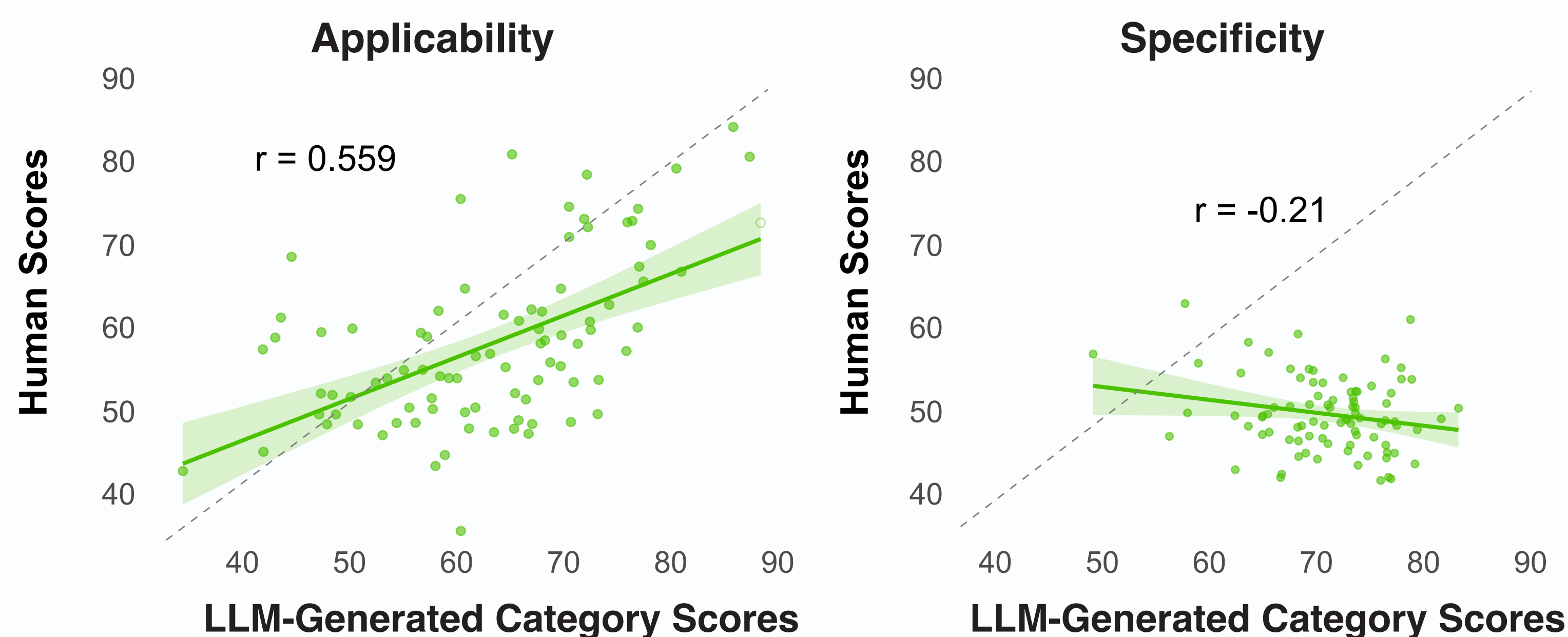


LLMs generate more “applicable” categories, especially for word sets composed of less similar words

Examples of Generated Categories

Generator	Word Set	Category
Human	case, death, party, couple	“All things that tend to appear in murder mysteries involving rich families”
Human	doctor, phone, hand, town	“Things that would be helpful in an apocalypse where the phone lines still work”
Human	murder, dude, afternoon, message	“Surfing a 1 PM when someone is killed and made to look like a suicide to send a message”
Model	heavens, stress, punishment, option	“Concepts often discussed in religious sermons about moral choices and their consequences”
Model	killer, surprise, space, idiot	“Sensational buzzwords often used in clickbait-style online titles (crime, shock, awe, and stupidity)”
Model	neck, arm, hat, computer	“Things you might adjust for comfort when settling in to use a computer (your neck position, arm position, hat, and the computer itself)”

Human/LLM Category Performance Correlations



The same word groups evoke similarly applicable but differently specific categories in LLMs and people.

References

Barsalou, L. W. (1983). Ad hoc categories. *Memory & cognition*, 11(3), 211-227.
Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38(11), 39–41.

Conclusions

LLMs generate both more specific and more applicable categories compared to people.

Semantically distant word groups prompt more specific LLM-generated and more general human-generated categories.

People likely settle on a “good enough” process of finding the conceptual glue.

Next Steps



What types of categories are rated as more specific and applicable?

colonel club agent wedding
“Things that can be seen in a mission impossible movie”

Why do some people consistently generate better categories?



Do people generate the same categories for the same word sets?