Virtual MathPsych/ICCM 2022
mathpsych.org
July 2022

Do models of syllogistic reasoning extend to generalized quantifiers?

Mr. Daniel Brand ~ Chemnitz University of Technology

Syllogisms

Some B are A.

All B are C.

What, if anything, follows?

- Two quantified statements that can be connected via a common term
- The task is to derive a conclusion about the two end terms or to conclude that nothing follows („no valid conclusion“, NVC)
- Well studied domain in reasoning research
- Most research is restricted to first-order logic quantifiers (All, None, Some, Some not)

Generalized Quantifiers

Most Mammals are Land Creatures.

Most Mammals are Intelligent Creatures.

What, if anything, follows?

- More quantifiers are important to understand **everyday reasoning**, as people use a variety of quantifiers: most, few, more than half,…
- Unknown how models for syllogistic reasoning perform on tasks with generalized quantifiers
- **However: Additional quantifiers drastically increase the number of tasks!**

Can models extend to generalized syllogisms?

- We extended the set of quantifiers to include **most** and **most not**
- Dataset: responses of 65 participants to all 144 tasks

- Evaluation of the **Probability Heuristics Model** (PHM) and **mReasoner**
- Models were fitted to **each individual participant** and then queried for a response to each task → **Accuracy**

mReasoner

- Based on the Mental Model Theorie (MMT)[1]
- Assumes that a mental model is constructed which represents the information of the premises via **instantiated sets**
- The conclusion (candidate) is derived based on the constructed model
- A **search for counterexamples** tests the conclusion candidates (can lead to NVC)
- Uses the ability to create **sets of differing sizes** to handle the generalized quantifiers

[1] e.g., Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, 107(43)

PHM

- Assumes that reasoners rely on a set of heuristics [1]
- Based on the informativeness of quantifiers:

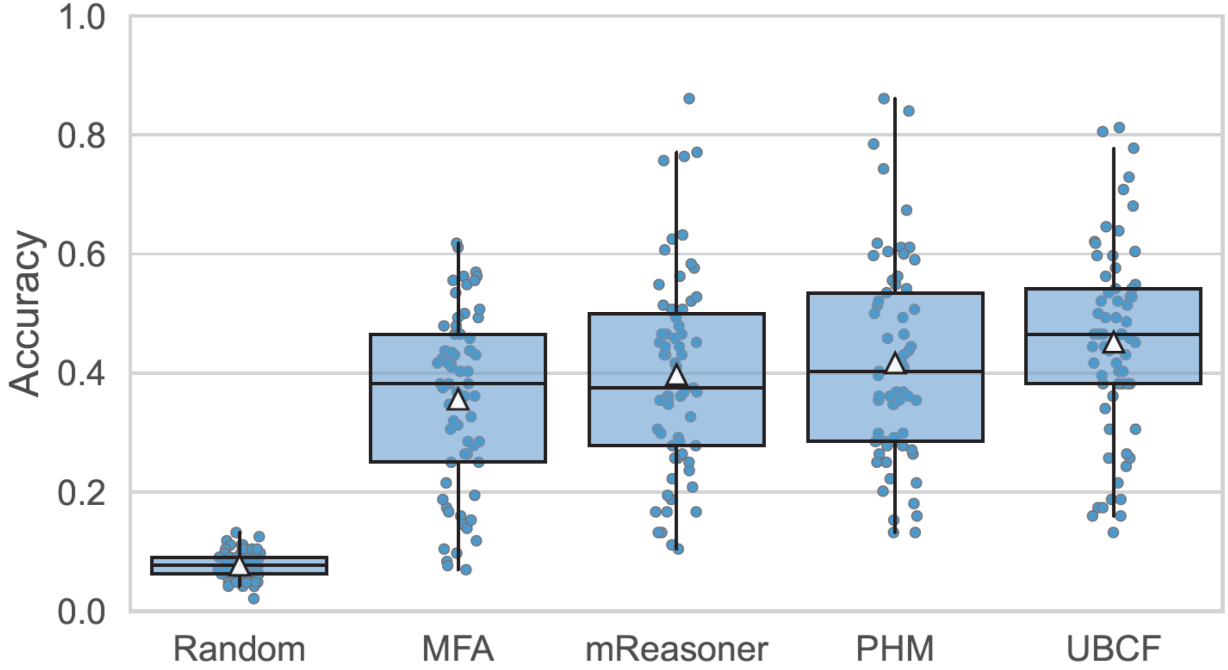
  All > Most > Most not > Some > No ≥ Some not
- Candidate conclusions are generated by 3 generative heuristics (*min-heuristic*, *p-entailment*, *attachement heuristic*)
- The candidates are then tested by 2 test-heuristics (*max-heuristic, O-heuristic*)
- Test heuristics are based on the „confidence“ in the quantifier (can lead to NVC)

[1] Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, 38(2), 191–258.

Baselines

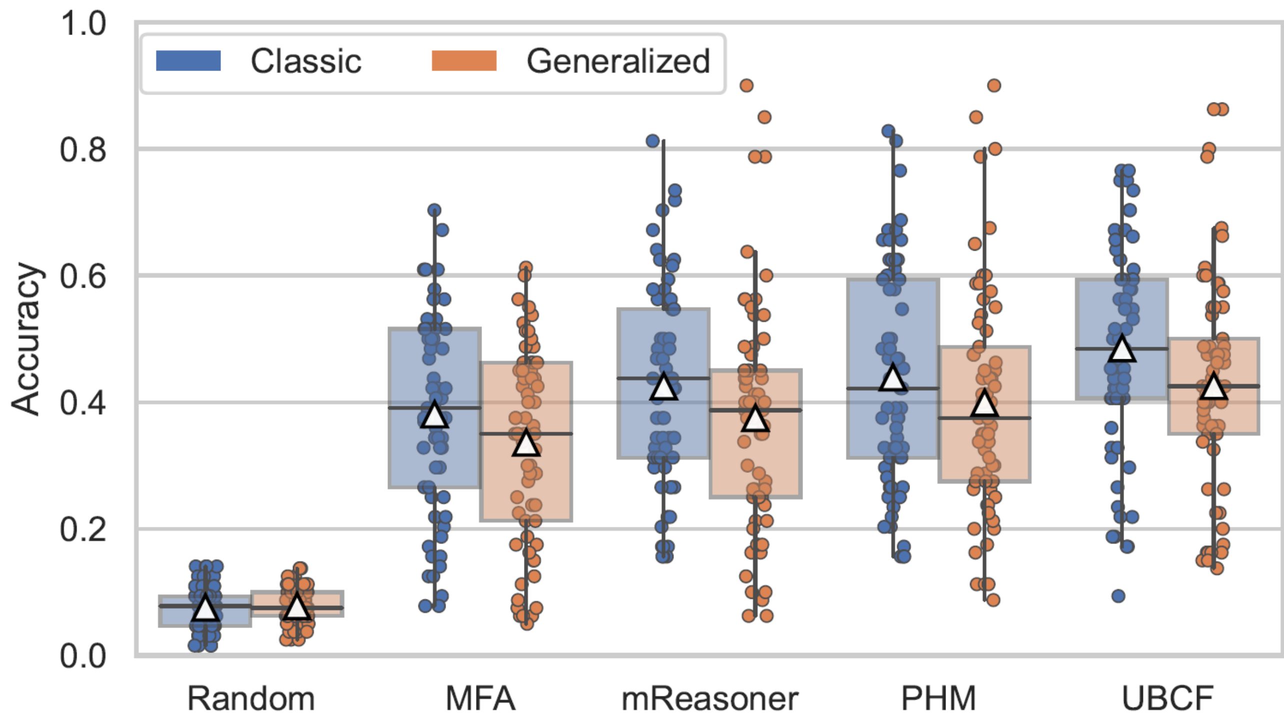
- The **most-frequently given answer** (MFA):
  - Upper bound for models that are not able to adapt to individuals
  - Expected lower bound for models that were fitted to individuals
- A **user-based collaborative filtering** model (UBCF):
  - Uses similar participants to determine a prediction for a task
  - Expected upper-bound for cognitive models (no restrictions due to limited parameters)

Results: Accuracy



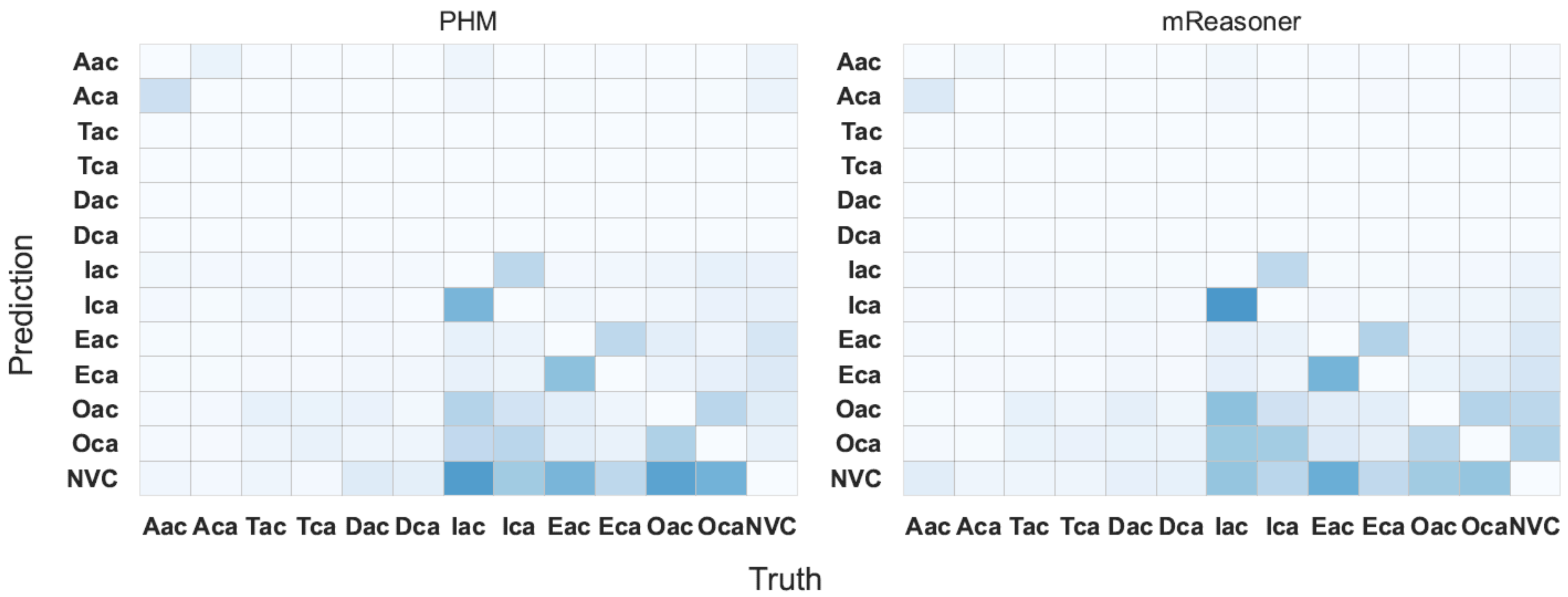
- Models were **not significantly better** than the MFA
- However, they seem to capture **some individuals** quite well

Results: Accuracy



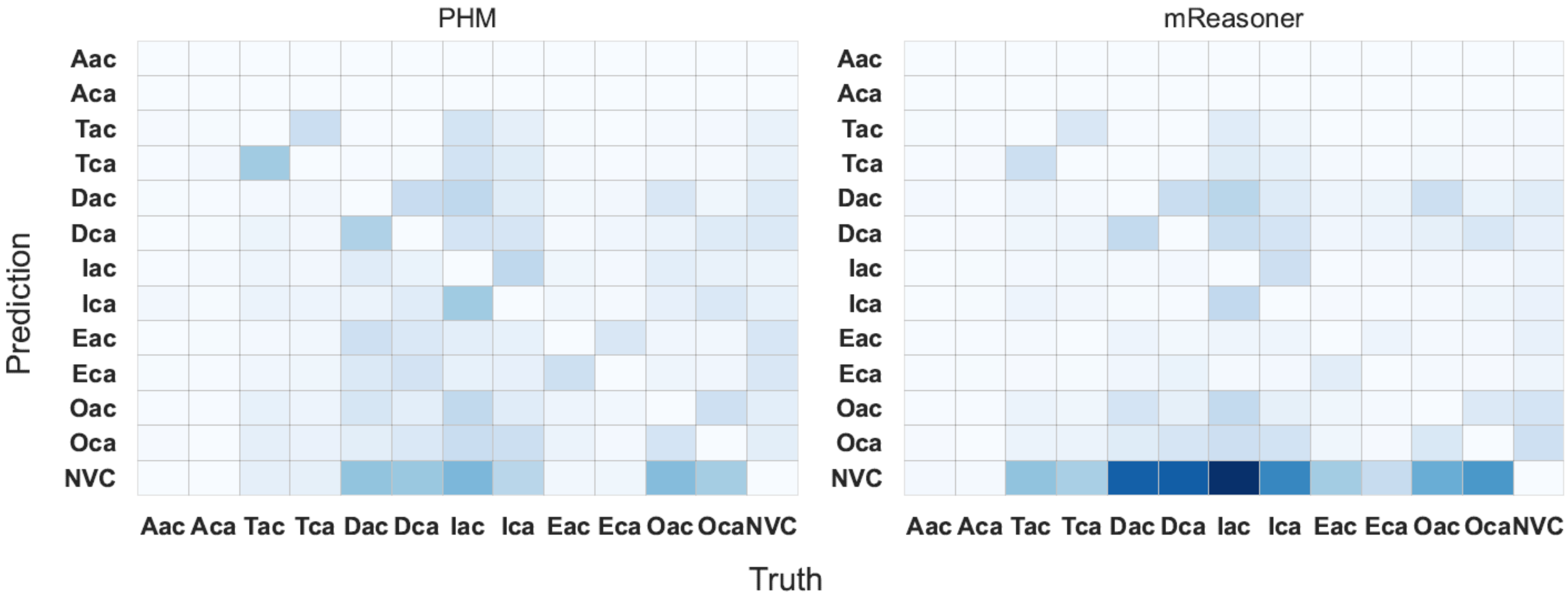
- Performance was worse on generalized quantifiers for all models
- Participants might behave more inconsistent in generalized tasks
- However: Where did the errors occur?

Source of error: Classic syllogisms



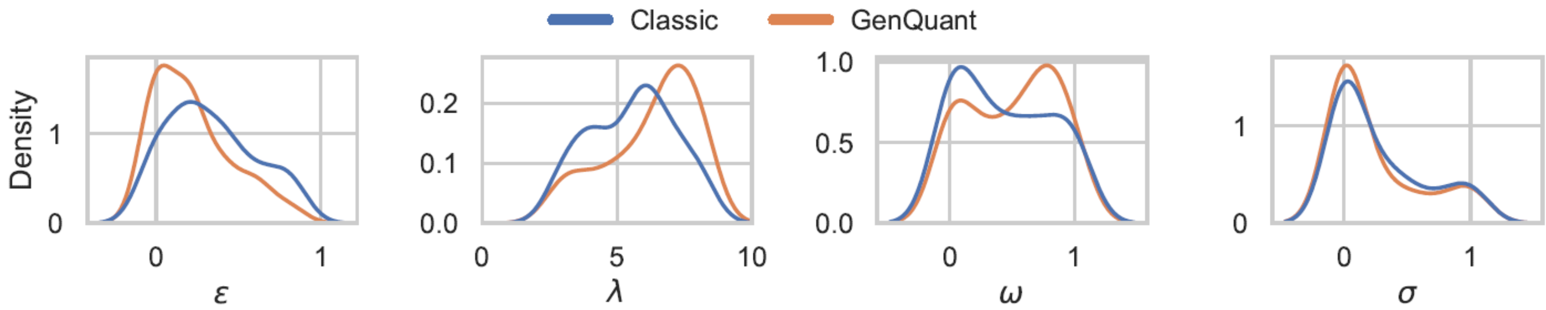
- NVC (models responding with NVC too often)
- Errors due to mixing up the directions: „Some A are C“ - „Some C are A“
- Errors for both models are similar

Source of error: Generalized quantifiers



- For PHM, the errors seem to be comparable to the classic syllogisms
- mReasoner's Errors were mostly due to NVC

mReasoner: Parameters



- λ controls the set sizes → difference is to be expected
- ω controls the **likelihood to continue** after a counterexample was found

→ High values for ω should lead to **low** NVC rates

Conclusions

- Models are generally able to extend to generalized tasks
- The performance of all models is substantially worse on generalized quantifiers
- PHM seems to be more stable with respect to the new quantifiers
  - However: It has additional parameters for additional quantifiers
- mReasoner differs greatly between generalized and classic tasks, showing that the processes are not yet unified
- NVC handling will be one of the most important aspects for generalized quantifiers (correct response to most tasks)
- It is essential to consider a larger set of quantifiers when modeling quantified reasoning processes

Thank you for your attention!