Probabilities in language and the world: Modeling syllogistic reasoning

Luuk Suurmeijer[1], Hana Filip[1], and Noortje J. Venhuizen[2]

[1]Institute for Language and Information, Heinrich Heine University Düsseldorf
[2]Department of Language Science and Technology, Saarland University

Syllogistic reasoning has a long history in the philosophical study of language, dating back to the work by Aristotle. Syllogisms are deductive arguments that consist of a set of premises and a conclusion. They have played a crucial role in formalizing logical deduction; in formal logic, syllogisms are classified as either valid or non-valid, where the former means that the conclusion is necessarily true given the conjunction of the premises. Example 1 displays a valid 'categorical' syllogism (a syllogism consisting of exactly three categorical propositions: two premises and one conclusion), which consists of three quantified propositions containing three distinct properties, each of which appears in two of the three propositions.

(1)
All singers are pianists.
Not all trumpet players are pianists.
___
Therefore: Not all trumpet players are singers

Example 1 is of the form *AOO-2*, as the first premise contains the quantifier 'all' (A), the second premise and the conclusion the quantifier 'not all' (O), and the order of the described properties (*singer*, *pianist*, *play trumpet*) is the second out of 4 possible 'figures', which define possible configurations of the properties. Based on 4 quantifiers (all, some, no, not all) and 4 figures, a total of 64 categorical syllogisms can be constructed, of which 27 constitute valid syllogisms.

Experimental work on human syllogistic reasoning has shown that humans do not adhere to a purely logical deductive process when tasked with evaluating the validity of a syllogism or with producing a conclusion based on the premises (Khemlani & Johnson-Laird, 2012). Humans seem to be on the one hand sensitive to the defeasibility and probability of the premises (Evans et al., 1983; Oakhill & Johnson-Laird, 1985; Stupple & Ball, 2008), but also to the ordering of the properties (figural effects) (Copeland & Radvansky, 2004; Jia et al., 2009; Stupple & Ball, 2007). Furthermore, human reasoners have difficulty identifying invalid syllogisms (Ragni et al., 2019; Riesterer et al., 2019) and exhibit large individual differences (Khemlani & Johnson-Laird, 2016). Here, we aim to arrive at a cognitive model for syllogistic reasoning that explains the pressures that underlie human syllogistic reasoning in terms of a trade-off between world knowledge and probabilistic language use.

In recent years, a number of models have been presented to explain human syllogistic reasoning, which can be roughly categorized into heuristic-based, logic-based and mental-models-based (Bischofberger & Ragni, 2020; Khemlani & Johnson-Laird, 2012). Critically, many of these models are defined verbally, and therefore do not allow for direct evaluation of their predictions. A notable exception is the recent model by Tessler et al. (2022), which is based on Rational Speech Act (RSA) theory (M. C. Frank & Goodman, 2012) and models syllogistic reasoning using probabilistic and pragmatic pressures on language comprehension and production. The model is evaluated based on the dataset from Ragni et al. (2019) and is shown to capture probabilistic effects on syllogistic reasoning by formalizing a reasoner's prior beliefs over possible situations. Here, we here aim to offer an explanation of human syllogistic reasoning at the algorithmic/representational level (in the sense of Marr, 1982) by modeling syllogistic reasoning in terms of incremental language comprehension. To this end, we employ the distributed meaning representations from Distributional Formal Semantics (DFS) (Venhuizen et al., 2021), a framework that models propositions as vectors of truth values over a set of formal model structures (effectively implementing a possible-world semantics perspective on propositions). The representations from DFS are inherently logical as well as probabilistic, which allows us in principle to capture the deductive as well as probabilistic aspects of syllogistic reasoning. Furthermore, DFS representations have been employed in recurrent neural networks and are shown to capture incremental and information-theoretic aspects of language comprehension (Venhuizen, Crocker et al., 2019a, 2019b; Venhuizen et al., 2021). In particular, Venhuizen, Crocker et al. (2019a) have shown that such models allow for quantifying incremental effects (in terms of Surprisal; cf. Levy, 2008) of independently manipulating world knowledge, modelled as the probabilistic variations in the DFS representations, and probabilistic language use, modelled as the distributional properties of the training corpus.

We model syllogistic patterns as the degree to which a conclusion can be inferred from the conjunction of the premises, where each statement is a DFS-vector. We can generate these vectors by sampling a vector space (meaning space) from a set of basic propositions and a set of constraints on model structures in DFS (see Venhuizen, Hendriks et al., 2019). The basic propositions are defined in terms of the properties a, b, c, applied to a set of 6 entities, resulting in a total of 18 propositions. Sampling then comes down to determining the truth values of each basic proposition in a large set of model structures ($N = 10000$) in accordance with the constraint set (an example constraint is that each entity in a model structure must have at least one property). The resulting set of models captures the world knowledge defined in the constraints as propositional co-occurrences across models: propositions that are true in many of the same models have related meanings. Furthermore, logical entailment is defined as subsumption across model structures: proposition p logically follows from proposition q *iff* p is true in all model structures in the meaning space that satisfy q. We can exploit the probabilistic nature of the meaning representations by using the inference score (see eq. (1) in the supplementary materials) (S. L. Frank et al., 2009) to compute to what extent meaning vectors are inferred from each other. Because DFS is compositional and all first order logic operations are defined over vectors, we can construct complex meaning vectors representing each of the three statements of a syllogism. Figure 1 shows the inference scores of all conclusions given the conjunction of the premises for a subset of the syllogisms. This figure shows that offline DFS representations capture full logical behavior, and moreover quantify uncertainty over possible conclusions.

We compare the inference scores from the 'offline' DFS representations to an existing dataset provided in the CCOBRA benchmarking package (Brand et al., 2019; see also Ragni et al., 2019). Given the inherently logical nature of these representations, they do not directly capture non-logical human reasoning behavior. Therefore, we exploit the probabilistic mapping from linguistic input to meaning that is offered by incorporating DFS representations in an RNN architecture. This allows us to manipulate the frequency with which premises occur in combination with each conclusion, which may be driven by, e.g., figural effects or the prior probability of the premises/conclusion. In addition, the incremental nature of such a model allows us to investigate the word-by-word inferences that are made during syllogistic comprehension. In sum, where previous work has illustrated the sensitivity of human syllogistic reasoning to broad communicative pressures (Tessler et al., 2022), our neural model using DFS representations offers a novel perspective on syllogistic reasoning in terms of incremental language comprehension: Reasoners use probabilistic properties of language and the world to determine valid inferences.

Supplementary Material

$$\inf(p,q) = \begin{cases} \frac{P(p|q)-P(p)}{1-P(p)} & P(p|q) > P(p) \\ \frac{P(p|q)-P(p)}{P(p)} & otherwise \end{cases} \tag{1}$$

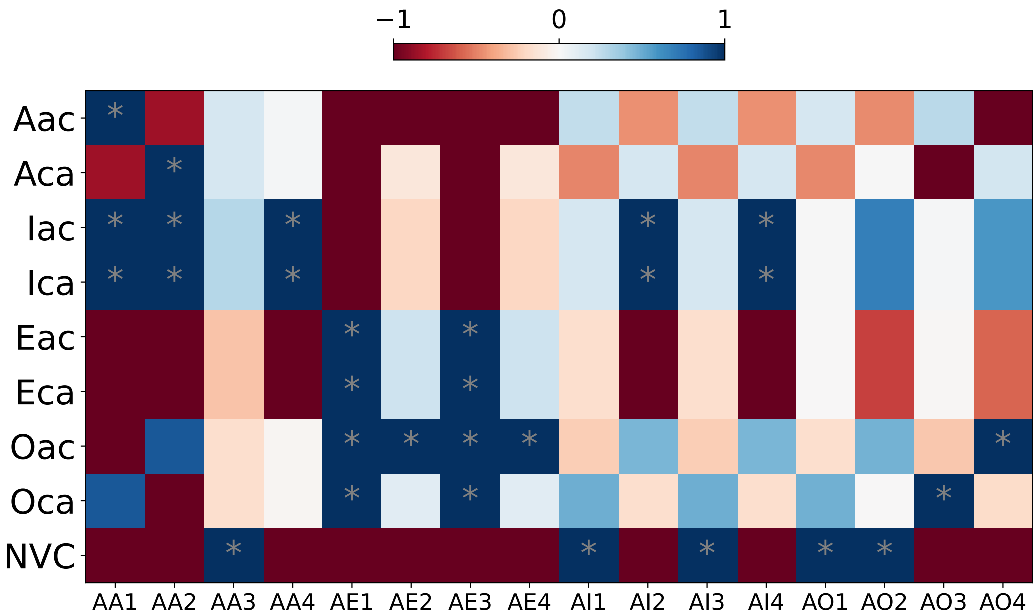

Figure 1: Inference scores between the premises of a subset of the 64 syllogisms and the 9 possible conclusions from a meaning space of 10000 models. Values of 1 (dark blue) and −1 (dark red) squares mean a positive and a negative entailment respectively. Real values in this range (shades squares) indicate an (un)likely conclusion. Logically valid conclusions are marked with a gray asterisk.

References

Bischofberger, J. & Ragni, M. (2020). Improving cognitive models for syllogistic reasoning. In S. Denison, M. Mack, Y. Xu & B. C. Armstrong (Eds.), *Proceedings of the 42th annual meeting of the cognitive science society - developing a mind: Learning in humans, animals, and machines, cogsci 2020, virtual, july 29 - august 1, 2020*. cognitivesciencesociety.org. https://cogsci.mindmodeling.org/2020/papers/0171/index.html

Brand, D., Riesterer, N. & Ragni, M. (2019). On the matter of aggregate models for syllogistic reasoning: A transitive set-based account for predicting the population.

Copeland, D. & Radvansky, G. (2004). Working memory and syllogistic reasoning. *The Quarterly journal of experimental psychology. A, Human experimental psychology*, *57*, 1437–57. https://doi.org/10.1080/02724980343000846

Evans, J. S. B. T., Barston, J. L. & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306. https://doi.org/10.3758/BF03196976

Frank, M. C. & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998. https://doi.org/10.1126/science.1218633

Frank, S. L., Haselager, W. F. G. & van Rooij, I. (2009). Connectionist semantic systematicity. *Cognition*, *110*(3), 358–379.

Jia, X., Lu, S., Zhong, N. & Yao, Y. (2009). Figural Effects in Syllogistic Reasoning with Evaluation Paradigm: An Eye-Movement Study. In N. Zhong, K. Li, S. Lu & L. Chen (Eds.), *Brain Informatics* (pp. 106–114). Springer Berlin Heidelberg.

Khemlani, S. & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, *138*(3), 427–457. https://doi.org/10.1037/a0026841

Khemlani, S. & Johnson-Laird, P. N. (2016). How people differ in syllogistic reasoning. *Cognitive Science*.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Marr, D. (1982). *Vision*. The MIT Press. https://doi.org/10.7551/mitpress/9780262514620.001.0001

Oakhill, J. V. & Johnson-Laird, P. N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *The Quarterly Journal of Experimental Psychology Section A*, *37*(4), 553–569. https://doi.org/10.1080/14640748508400919

Ragni, M., Dames, H., Brand, D. & Riesterer, N. (2019). When does a reasoner respond: Nothing follows? In A. K. Goel, C. M. Seifert & C. Freksa (Eds.), *Proceedings of the 41th annual meeting of the cognitive science society, cogsci 2019: Creativity + cognition + computation, montreal, canada, july 24-27, 2019* (pp. 2640–2546). cognitivesciencesociety.org. https://mindmodeling.org/cogsci2019/papers/0452/index.html

Riesterer, N., Brand, D., Dames, H. & Ragni, M. (2019). Modeling human syllogistic reasoning: The role of "no valid conclusion". In A. K. Goel, C. M. Seifert & C. Freksa (Eds.), *Proceedings of the 41th annual meeting of the cognitive science society, cogsci 2019:*

Creativity + cognition + computation, montreal, canada, july 24-27, 2019 (pp. 953–959). cognitivesciencesociety.org. https://mindmodeling.org/cogsci2019/papers/0179/index.html

Stupple, E. J. N. & Ball, L. (2007). Figural Effects in a Syllogistic Evaluation Paradigm: An Inspection-Time Analysis. *Experimental psychology*, *54*, 120–7. https://doi.org/10.1027/1618-3169.54.2.120

Stupple, E. J. N. & Ball, L. J. (2008). Belief–logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning*, *14*, 168–181.

Tessler, M. H., Tenenbaum, J. B. & Goodman, N. D. (2022). Logic, probability, and pragmatics in syllogistic reasoning. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12593

Venhuizen, N. J., Crocker, M. W. & Brouwer, H. (2019a). Expectation-based comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes*, *56*(3), 229–255. https://doi.org/10.1080/0163853x.2018.1448677

Venhuizen, N. J., Crocker, M. W. & Brouwer, H. (2019b). Semantic entropy in language comprehension. *Entropy*, *21*(12), 1159. https://doi.org/10.3390/e21121159

Venhuizen, N. J., Hendriks, P., Crocker, M. W. & Brouwer, H. (2019). A framework for distributional formal semantics. In R. Iemhoff, M. Moortgat & R. de Queiroz (Eds.), *Logic, language, information, and computation* (pp. 633–646). Springer. https://doi.org/10.1007/978-3-662-59533-6_39

Venhuizen, N. J., Hendriks, P., Crocker, M. W. & Brouwer, H. (2021). Distributional formal semantics. *Information and Computation*, 104763. https://doi.org/https://doi.org/10.1016/j.ic.2021.104763