# Modeling Human Syllogistic Reasoning: The Role of "No Valid Conclusion"

**Nicolas Riesterer* (riestern@cs.uni-freiburg.de)**

**Daniel Brand* (daniel.brand@cognition.uni-freiburg.de)**

**Hannah Dames (damesh@cs.uni-freiburg.de)**

**Marco Ragni (ragni@cs.uni-freiburg.de)**
Cognitive Computation Lab, Georges-Köhler-Allee 79
79110 Freiburg, Germany

**Abstract**

"No Valid Conclusion" (NVC) is one of the most frequently selected responses in syllogistic reasoning experiments and corresponds to the logically correct conclusion for 58% of the syllogistic problem domain. Still, NVC is often neglected in computational models or just treated as a byproduct of the underlying inferential mechanisms such as a last resort when the search for alternatives is exhausted. We illustrate that NVC represents a major shortcoming of current models for human syllogistic reasoning. By introducing heuristic rules, we demonstrate that slight extensions of the existing models result in substantial improvements of their predictive performances. Our results illustrate the need for better NVC handling in cognitive modeling and provide directions for modelers on how to integrate it into their approaches.

**Keywords:** cognitive modeling; heuristics; syllogistic reasoning; no valid conclusion

## Introduction

Syllogistic reasoning is one of the core domains in human reasoning research (for a review see Khemlani & Johnson-Laird, 2012). It is concerned with gaining insight into the cognitive processes driving the inference mechanisms for categorical assertions featuring quantifiers ("All", "Some", "Some ... not", and "No") and terms which are inter-related by two premises. The traditional experimental paradigm presents participants with problems of the form "All A are B; All B are C" (substituting A, B, and C with common groups such as gardeners, musicians, etc.) and usually asks "What follows?", i.e., which conclusion can be inferred logically from the premises (generation task; Morley, Evans, & Handley, 2004). Depending on the arrangement of terms, the syllogism is categorized into one of four figures, a property that was found to have a substantial influence on human inferences (Johnson-Laird & Bara, 1984):

| | Figure 1 | Figure 2 | Figure 3 | Figure 4 |
|---|---|---|---|---|
| Premise 1 | A-B | B-A | A-B | B-A |
| Premise 2 | B-C | C-B | C-B | B-C |

For reasons of clarity, syllogistic problems are usually referred to by abbreviating quantifiers with single uppercase letters and the figure number: "All" (A), "Some" (I), "No" (E), "Some ... not" (O). The syllogism "All informative things are useful; Some websites are not informative things" is therefore referred to as AO2. Possible conclusions for syllogistic problems combine the end terms A and C via one of the four quantifiers. Additionally, it is possible to respond with "No Valid Conclusion" (NVC) indicating that the premises have no valid conclusion in accordance to first-order logic. Out of the 64 distinct syllogistic problems, 37 are invalid (58%), i.e., only NVC can be derived.

Experimental investigations have shown that NVC represents one of the most frequently selected conclusions (Khemlani & Johnson-Laird, 2012). Because of this, the role of NVC in syllogistic reasoning is important. However, current models of syllogistic reasoning rarely make explicit statements about NVC. On the extreme, there are heuristic models which do not possess the capability of generating NVC at all. On the other hand, models that do integrate NVC as a conclusion candidate often treat it as a termination criterion when searches for alternatives fail. Currently, there are no strategies to directly infer NVC responses. Additionally, even when going beyond the level of predictions, models are unable to account for statistical phenomena related to NVC responses, such as variations in reaction times (Ragni, Dames, Brand, & Riesterer, 2019).

In this article, we tackle this problem by proposing a set of heuristic rules for generating NVC conclusions based on findings from the syllogistic literature. By attaching these rules to existing models, we show that inadequate NVC handling is indeed one of the core problems of the current state of the art. The following text is split into five sections. After introducing the syllogistic domain of reasoning as well as the current state of the art in modeling (Section 2), we will analyze contemporary models in terms of their capabilities in predicting a human NVC response (Section 3). Section 4 then takes up those results and presents alternative strategies for predicting NVC responses. In Section 5 we evaluate the syllogistic models augmented with the identified strategies for NVC and finally, in Section 6, discuss our results, illustrate the potential with respect to improving models, and give directions for future work in the field of cognitive modeling of human reasoning.

*Both authors contributed equally to this manuscript.

Table 1: Models and their NVC prediction proportions for valid and invalid syllogisms. For models with multiple prediction candidates for a single syllogism, the ratio of NVC is used. Model predictions are taken from Khemlani and Johnson-Laird (2012).

| Model | Valid | Invalid |
|---|---|---|
| Conversion | 44% | 86% |
| Mental Models Theory (MMT) | 14% | 30% |
| PSYCOP | 0% | 100% |
| Verbal Models | 32% | 51% |
| Atmosphere | 0% | 0% |
| Matching | 0% | 0% |
| Probability Heuristics Model (PHM) | 0% | 0% |

## Related Work

Computational modeling is a central part of today's research of human syllogistic reasoning. As of today, at least twelve theories about syllogistic inferences exist. In a meta-analysis, Khemlani and Johnson-Laird (2012) found that the theories have distinct advantages and drawbacks when predicting experimental data obtained by aggregating individual participants' responses. The following paragraphs briefly introduce the different approaches for which the authors were able to provide predictions for the 64 syllogisms. They will be used throughout the following analyses.

The **Conversion Hypothesis** is an attempt at explaining erroneous conclusions resulting from human reasoning processes originally introduced by Chapman and Chapman (1959) and later formalized as a testable model by Revlis (1975). The hypothesis states that while encoding a syllogistic premise, a conversion operation is applied which swaps the direction of the categorical expression (e.g., "All A are B" is interpreted as "All B are A"). As a result, a new syllogism is produced with conclusions that might be inappropriate for the original problem (e.g., Revlin, Leirer, Yopp, & Yopp, 1980). NVC is predicted if the new problem is logically invalid.

The **Mental Model Theory** (MMT; Johnson-Laird, 1975) is a cognitive theory which has successfully been applied to various domains of reasoning (Johnson-Laird & Byrne, 2002; Khemlani & Johnson-Laird, 2012; Ragni & Knauff, 2013). It is based on the assumption that inferential mechanisms operate on mental representations constructed for the given premises. MMT's inference process is composed of a series of phases: model construction, conclusion generation, and the search for counterexamples. First, an initial mental model is constructed integrating the information of the premises, i.e., the relation between the terms of the premises. Second, a candidate conclusion is formulated in accordance to the initial model. Finally, alternative models consistent to the premises are constructed in search of a situation in which the conclusion is false (Ragni, Khemlani, & Johnson-Laird, 2014). If the initial model construction fails, or counterexamples can be found for all models, NVC is returned.

The **Psychology of Proof** model (PSYCOP; Rips, 1994) is a cognitive model of human syllogistic reasoning that claims deduction as a fundamentally human capability (Khemlani & Johnson-Laird, 2012). PSYCOP defines a set of psychologically plausible inference rules approximating the human inferential mechanisms. By applying rules in a deductive forward-inference fashion as well as an inductive backwards-inference fashion, a path between premise information and conclusion is constructed. PSYCOP does not have a guaranteed way to conclude NVC. While it supports exhaustive searches for conclusions and the generation of NVC as fallback option, this behavior is not enforced in its original formulation (Khemlani & Johnson-Laird, 2012).

**Verbal Reasoner** (Polk & Newell, 1995) is an approach to modeling syllogistic reasoning that assumes that human inferences are fundamentally verbal. It encodes the premise information into a mental model that differentiates between more accessible information (the subject of the premise) and less accessible information (the object of the premise). By defining procedures to extract different degrees of intermediate implicit knowledge about the reasoning problem, the model is able to generate conclusions following more or less complex inferences. The verbal model theory treats NVC as a last-resort option. If no conclusion can be derived from the mental model, the verbal reasoner enters a reencoding loop in search for a solution. NVC is produced when it gives up.

The **Atmosphere Hypothesis** (Woodworth & Sells, 1935) is able to account for a portion of errors in human syllogistic reasoning when compared with formal logics (Revlis, 1975). It is based on a feature extraction step that identifies whether the given premise information is positive/negative ("All", "Some" vs. "Some not", "No") and universal/particular ("All", "No" vs. "Some", "Some not"). By following a combination procedure, the quantifier of the conclusion is determined. Because it only extracts and combines features based on quantifiers, the atmosphere hypothesis is not able to provide information about the direction, i.e., the order of terms in the syllogistic conclusion, and is not able to generate NVC.

The **Matching Hypothesis** (Wetherick & Gilhooly, 1995) reflects a different approach for accounting for errors made in human syllogistic reasoning. It employs a matching strategy which states that the conclusion quantifier is equal to the most conservative quantifier in the premises. Conservativeness in this sense is defined as a preference order of $E > O = I \gg A$ following the estimated number of individuals a quantifier makes a statement about. Similar to Atmosphere, Matching is unable to predict NVC, because it always picks a quantifier from the given premises.

The **Probability Heuristics Model** (PHM; Chater & Oaksford, 1999) is an approach to modeling reasoning that is based on the fundamental idea that reasoning relies on heuristics. PHM defines the inferential process via two phases. First, a conclusion is generated by applying the *min-heuristic* selecting the least informative quantifier from the premises ($A > I > E \gg O$). Second, *probabilistic entailments* can be

applied generating alternative conclusions based on the min-heuristic's result that could probably be true. Next, a third heuristic, *attachment*, is applied to determine the order of terms in the conclusion. Finally, the *max-heuristic* is applied to assess the confidence of the conclusion based on the informativeness of the premises. If confidence is low, the probability of returning NVC instead of the solution candidate rises. Additionally, the *o-heuristic* is applied which states that O-responses should be avoided in favor of NVC.

In Khemlani and Johnson-Laird (2012)'s prediction table, which we use as the source for the models' predictions, PHM is reported without an inclusion of the max- and o-heuristic (Baratgin et al., 2015). While potentially distorting for model comparisons, this does not affect our evaluation of NVC. The max- and o-heuristics are attached to PHM's inference mechanisms (min-heuristic, attachment, and probabilistic entailment) in similar spirit to what we propose as general extensions of cognitive models further below.

The present article investigates the theories based on their NVC prediction capabilities. Table 1 summarizes the models' NVC response proportions in accordance to the prediction data reported by Khemlani and Johnson-Laird (2012) for valid and invalid syllogisms. The table highlights the difference between the cognitive models. While some models are unable to predict NVC at all, the other approaches have a stronger tendency toward responding with NVC for invalid syllogisms. This behavior is expected due to NVC being the logically valid response for invalid syllogisms. PSYCOP reflects formal first order logic in its NVC response behavior. Because all valid and no invalid syllogisms have categorical conclusions, it predicts 0% and 100% NVC, respectively. In the following analyses, we evaluate the models based on their ability to predict the most frequently selected responses.

## Analysis State of the Art

### Modeling Task

In this article, we aim at uncovering the latent potential of the current state of the art by investigating their prediction capabilities with a special focus on NVC. Hence, we adopt a predictive scenario as the core evaluation setting of the following analyses: Given a dataset of reasoning data, we first compute the most frequent answer (MFA) and assess each model's performance by comparing its predictions with the aggregated response given by the participants.

The dataset used for this article was recorded as an Amazon Mechanical Turk web experiment in 2016 and consists of $N = 139$ participants providing conclusions to all 64 syllogistic problems, each. Participants were asked to select one of the nine syllogistic response candidates following from the premises. After a training phase consisting of four easy syllogisms, the remaining task sequence and order of response options was fully randomized.

The predictions for the model candidates were taken from Khemlani and Johnson-Laird (2012). This prediction data does not feature single explicit conclusions for each model and task. Instead, only sets of possible conclusions can be provided for each model and syllogism. To account for this in our prediction setting, weighted scores were computed for the following analyses via $S(P,T) = |P \cap T|/|P|$, where $P$ and $T$ denote the sets for predicted and true responses, respectively (e.g., Copeland, 2006).

All materials used for the following analyses are openly available via Github[1].

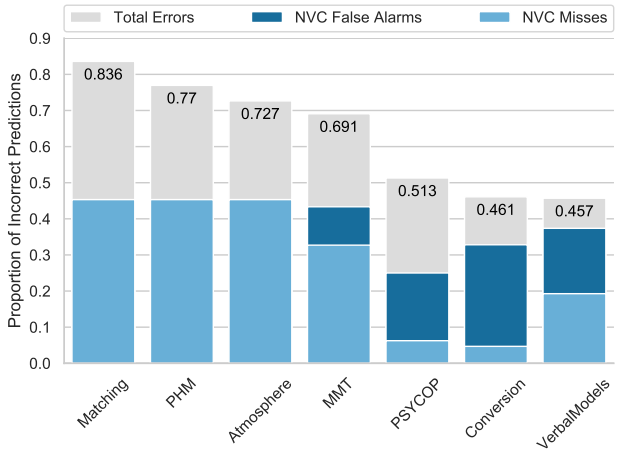


Figure 1: Proportion of model prediction errors (grey) for the 64 syllogisms. False alarms (dark blue), i.e., incorrect, and missed NVC predictions (light blue) are represented as proportions of the prediction error.

### State of the Art

Figure 1 illustrates the predictive capabilities of the models in accordance to the prediction table of Khemlani and Johnson-Laird (2012). The grey bars reflect the proportion of incorrect predictions on the 64 syllogisms' MFA responses. Dark blue and light blue bars denote the parts of incorrect responses which can be attributed to unwanted and missed NVC responses, respectively. As an illustrating example, PSYCOP incorrectly predicts 51% of the syllogisms. About 6% of those errors can be attributed to missed NVC responses whereas 19% of the errors were due to false alarms.

The plot highlights the difference between the models in today's state of the art. As expected, the models which are unable to predict NVC responses (Matching, PHM, Atmosphere), perform worst. For the remaining models, the general performance is better. However, NVC-based errors still account for the large parts of the incorrect predictions. As a particularly striking example, more than half of Conversion's errors are due to incorrect NVC predictions.

The depicted results highlight the need for a better understanding of NVC. In the following, we propose strategies for predicting NVC based on results from the literature on human syllogistic reasoning. Since embedding these strategies

[1] https://github.com/nriesterer/syllogistic-nvc

Table 2: Change in predictive accuracy (black, first value), misses (lightblue, middle value), and false alarms (darkblue, right value) of the models' NVC predictions.

| Models | PartNeg | | | EmptyStart | | | FiguralRule | | | NegativityRule | | | ParticularityRule | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Atmosphere | **37.5**% | -28 | 8 | 7.8% | -6 | 2 | 14.8% | -17 | 15 | 25.0% | -16 | 0 | 15.6% | -12 | 4 |
| Conversion | **1.6**% | -2 | 2 | 0.8% | -1 | 1 | -3.1% | -2 | 8 | 0.0% | 0 | 0 | 0.0% | 0 | 0 |
| MMT | **28.3**% | -20.1 | 6 | 6.2% | -4.4 | 1.6 | 11.3% | -12.3 | 12.2 | 17.5% | -11.2 | 0 | 11.3% | -8.4 | 2.8 |
| Matching | **42.2**% | -28 | 8 | 7.8% | -6 | 2 | 19.5% | -17 | 15 | 25.0% | -16 | 0 | 17.2% | -12 | 4 |
| PHM | **39.8**% | -28 | 8 | 7.8% | -6 | 2 | 17.6% | -17 | 15 | 25.0% | -16 | 0 | 16.4% | -12 | 4 |
| PSYCOP | **4.2**% | -4 | 4 | 3.1% | -2 | 2 | -3.0% | -2 | 13 | 0.0% | 0 | 0 | 0.0% | 0 | 0 |
| VerbalModels | **11.6**% | -11.7 | 5.2 | 4.2% | -3.2 | 1 | 3.3% | -6.7 | 8.4 | 9.1% | -5.8 | 0 | 4.9% | -4.7 | 2 |

into the assumptions stemming from the high-level theoretical ideas of the models exceeds the scope of this article, we focus on formulating the NVC strategies as rules which can be attached to arbitrary models. If a rule does not predict NVC, the underlying model is queried. This allows us to examine the benefits and assess potential shortcomings of an improved NVC handling in modeling human syllogistic reasoning. Because our rules are purely additive, we expect models with high numbers of NVC misses to benefit most from the proposed strategies. The challenge lies in minimizing the inevitable increase in false alarms.

## Towards a Model of NVC

To tackle the problem of missed NVC responses, we introduce a set of heuristic rules detecting NVC which are based on different observations.

The first heuristic, the *Figural Rule* is based on the figural effect, a core result of syllogistic reasoning research. Early studies found that the figure of premises induces a reliable bias on participants' responses: Figure 1 encourages A-C responses while Figure 2 leads to higher proportions of C-A responses (Johnson-Laird, 1975). In a later study it was found that the syllogistic figure also has an effect on the proportion of NVC responses (Johnson-Laird & Bara, 1984): NVC is preferred for syllogisms of Figure 3 and 4. This finding is transformed into a rule generating the NVC response whenever a syllogism of Figure 3 and 4 is encountered. For the remaining figures, the attached model is queried.

The next set of rules draws from the notion of informativeness of quantifiers as a criterion for determining NVC. Informativeness is a driving factor for two models in the current state of the art of syllogistic reasoning. The probability heuristics model (Chater & Oaksford, 1999) assumes an informativeness ordering of $A > I > E \gg O$ based on how unexpected truth about a statement is conceived by humans. Matching, on the other hand, introduces the notion of conservativeness based on the number of individuals a premise makes an assertion about: $E > O = I \gg A$ (Wetherick & Gilhooly, 1995). Both orders assign the least amount of information to the negative quantifiers "Some ... not" (O), and "No" (E). The *negativity rule* integrates both orders by being defined on the assumption that the amount of information encoded by two negative premises does not suffice to license a valid conclusion. This rule relates to PHM's max-heuristic in the sense that it assumes a threshold for insecurity with a generated conclusion candidate that is exceeded for E and O quantifiers. In doing so, it also subsumes PHM's o-heuristic. In analogy to negativity, the *particularity rule* is defined based on the limited information encoded in the particular quantifiers "Some" (I) and "Some ... not" (O). They make assumptions about limited and unspecified sets which might cause the reasoning process to fail. Finally, we define a third rule, *PartNeg* by combining both particularity and negativity: If the syllogism only consists of quantifiers with limited information, i.e., does not contain "All", NVC is predicted.

The last rule, *EmptyStart*, focuses on the syllogisms where information can be propagated transitively through the premises. This is possible for figure 1, i.e., "A-B, B-C", or figure 2, i.e., "B-A, C-B", which can be converted into figure 1 by swapping the premises and substituting C with A and A with C. The heuristic assumes that an information propagation is constructed (A-B-C for figure 1, C-B-A for figure 2). Inferences can only be drawn if the quantifier relating the two terms in the beginning of the chain makes an assertion about a non-empty set of individuals. If this premise features "No", i.e., the most conservative premise (Wetherick & Gilhooly, 1995), no information can be propagated through the chain and NVC is inferred. If we consider syllogism IE1, the chain A-B-C can be extracted starting with quantifier "Some". The reasoner is able to identify a selection of elements from A which can be annotated as B. The information from the second premise can now be integrated easily into the elements from A. If we consider EI1 on the other hand, the reasoner is unable to identify an initial set of elements from A. Therefore, premise 2 cannot be related to elements from A. As a result, there is a higher chance to respond with NVC.

### Analysis

Figure 2 depicts the syllogisms for which the introduced heuristics predict NVC along with the syllogisms for which NVC is the most frequent answer (MFA). Comparing the strategies our results show that different parts of the space of syllogisms are covered by different rules. For instance,

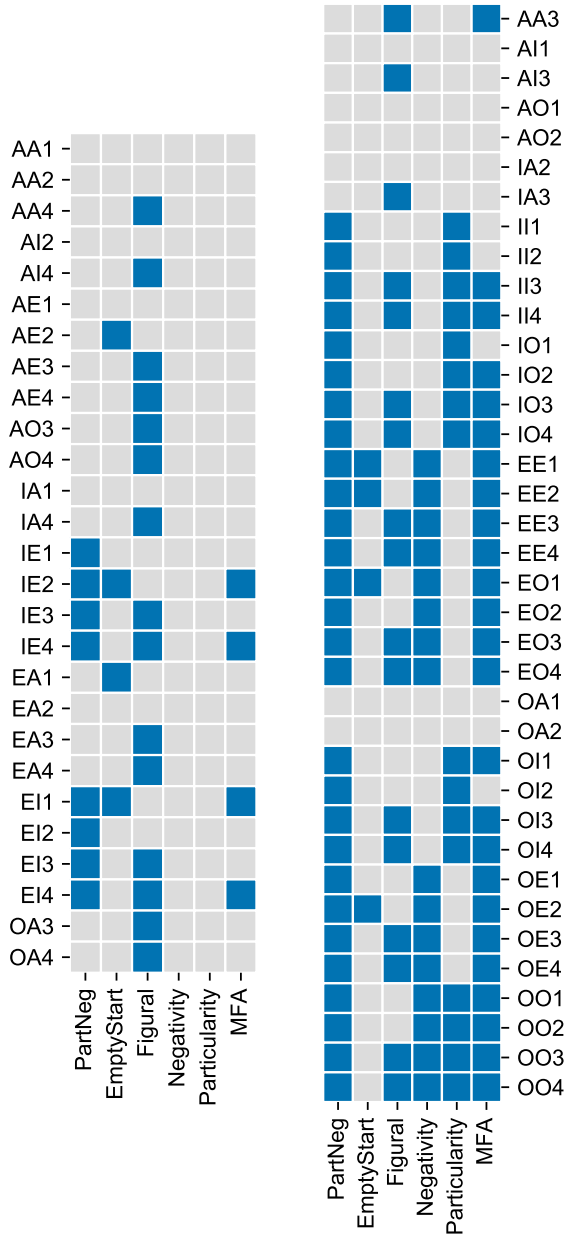


Figure 2: NVC Predictions of the individual rules on valid (left) and invalid syllogisms (right). Syllogisms are abbreviated with the encoded quantifiers of both premises "All" (A), "Some" (I), "No" (E), and "Some ... not" (O), and the figure.

negativity and particularity do not predict NVC for valid syllogisms, because there only exist invalid syllogisms characterized by being fully negative or particular. Figural on the other hand generates NVC for large parts of the syllogistic domain regardless of the validity of the underlying problem. When compared with MFA, the rules vary in predictive performance. PartNeg is capable of covering large parts of the invalid syllogisms correctly and only makes few errors for the valid cases. In contrast, figural's predictions show a more substantial difference in performance between valid and invalid syllogisms.

More generally, the plot also illustrates that most responses were not given by following standard logics. This is especially apparent in the case of the 37 invalid syllogisms where only 25 (68%) of the MFA responses correspond to NVC.

## Integrating NVC into Models

To determine the effectiveness of our NVC rules, we attach them to the original state-of-the-art models and evaluate their change in performance. This is depicted in Table 2. It presents the raw improvement of the syllogistic models achieved by attaching the respective NVC rule. Additionally, the decrease in misses (light blue) and increase in false alarms (dark blue) are illustrated. In general, larger improvements (percentages), fewer misses, and fewer false alarms indicate better performance.

Table 2 draws a convincing picture about the qualities of the NVC rules. With the exception of the figural rule, all strategies result in substantial improvements over the standard models. PartNeg achieves the overall peak performance improving up to 42.2% when compared to the base model. EmptyStart has the overall lowest changes in performance but introduces only few additional errors. As expected, models which do not generate NVC at all benefit most from the capability of responding with NVC achieving an improvement of 21.3%, 20.3%, and 20.1% on average across all NVC rules, respectively. PSYCOP (0.9% on average) and Conversion (-1.4% on average) do not benefit from the additional NVC rules with Conversion's performance even decreasing slightly. Surprisingly though, MMT is improved substantially by the additional NVC rules (14.9% on average) even though it already has the capability of generating NVC.

To gain additional insight into the performance of the models, Figure 3 replicates the introductory plot from Figure 1. It depicts the errors in the predictions of the models extended with PartNeg, the overall best NVC rule. Again, the plot depicts the proportion of incorrect predictions (grey) as well as the fractions corresponding to false alarms (dark blue) and misses (light blue).

The figure illustrates that the attached rule, PartNeg, manages to effectively remove NVC misses from the models' predictions. Simultaneously, it achieves this without introducing substantial amounts of false alarms. Consequently, in combination with PartNeg, a heuristic rule was found that is able to nuancedly relate human reasoner's tendencies towards concluding NVC to the syllogistic quantifiers. The fact that the improvement in handling NVC caused a substantial increase in performance for most of the models further strengthens the claim that NVC is one of the core weaknesses of the current state of the art in modeling human syllogistic reasoning.

Figure 4 illuminates the qualities of NVC rules on an individual level. For each model, the values refer to the number of participants for which a certain rule achieves highest

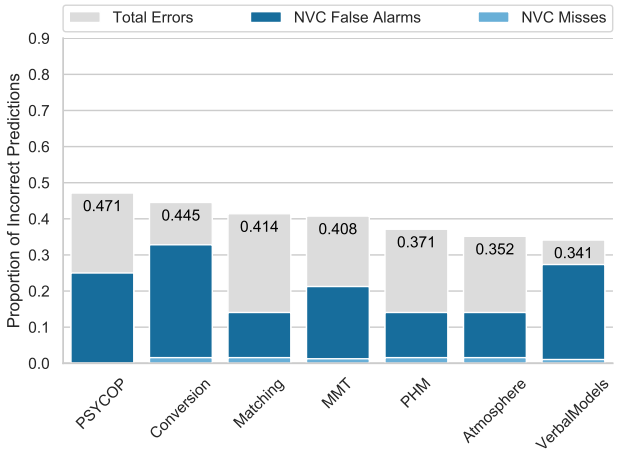


Figure 3: Proportion of the prediction errors (grey) achieved by the models extended with PartNeg, the overall best NVC heuristic, on the 64 syllogisms. False alarms (dark blue), i.e., incorrect, and missed NVC predictions (light blue) are represented as proportions of the prediction error.

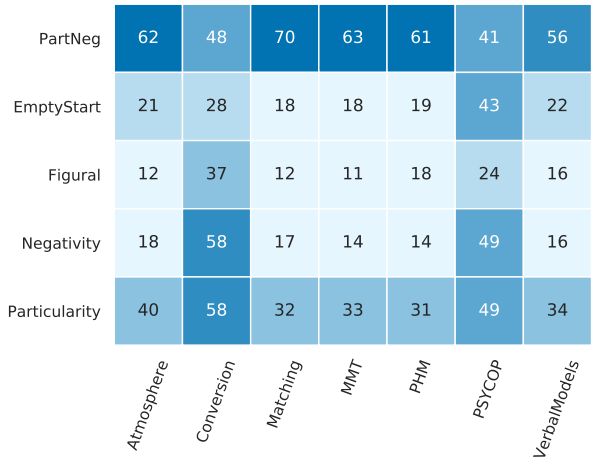


Figure 4: For each model, the values denote the number of participants for which the corresponding NVC rules performs best. In case of ties, the subject is counted for both rules.

performance. The figure illustrates that while PartNeg is the overall best rule, there is quite a substantial number of participants which can be accounted for better by other rules. This suggests that NVC response behavior is dependent on inter-individual differences of reasoning processes.

## General Discussion

As the correct response for 58% of syllogisms as well as one of the most frequently given responses by human reasoners (Khemlani & Johnson-Laird, 2012), "No Valid Conclusion" (NVC) is an important response for computational models to capture.

Our results demonstrate that the current state of the art in modeling human syllogistic reasoning is lacking the capabilities for handling NVC correctly. While some other approaches do not feature the ability of producing NVC at all, even the more complex approaches yield false alarm rates of up to 25% (Conversion) and misses of up to 30%. The high miss rates highlight a lack of precision in identifying the problems where NVC responses are adequate.

We combat these shortcomings by introducing five heuristic rules for predicting NVC based on prominent phenomena and properties of syllogistic reasoning (e.g., figural effect; Johnson-Laird, 1975; Johnson-Laird & Bara, 1984, or informativeness of premises; Chater & Oaksford, 1999). By attaching these rules to the cognitive models taken from Khemlani and Johnson-Laird (2012), a substantial improvement can be observed for the majority of models. Models without the capability of predicting NVC could achieve an increase in performance of up to 20% on average across all rules. Combined with PartNeg, the overall best NVC rule, we were able to demonstrate a substantial decrease of misses across the board. Even though these rules introduce low numbers of additional false alarms, this effect is negligible when compared to the substantial reduction of misses.

In conclusion, our work contributes to research in the domain of syllogistic reasoning both on a theoretical and practical level. We isolate NVC as one of the core flaws of the current state of the art in modeling syllogistic reasoning. By demonstrating substantial improvement when attaching NVC predictors, we highlight the remaining potential for modelers to tap into. The next step for cognitive modelers is to integrate these findings into future iterations of their models and derive additional rules from cognitive theories. With PartNeg, we provide a first rule which represents a valuable heuristic candidate for explaining NVC response behavior.

Furthermore, our results show the potential that lies in isolating and improving parts of the problem domain. By highlighting their shortcomings, modelers are given the chance to iteratively improve on their computational models and underlying theories. Apart from NVC, another candidate for improvement is the conclusion direction. Currently, there exist models which completely ignore direction as a predictive factor (e.g., Atmosphere) and others which actively integrate it into their underlying formalisms (e.g., Conversion).

Still, even though PartNeg captures the majority of MFA responses, it is not the optimal choice for each individual. There still is potential left for making better predictions if the relation between individual reasoners' characteristics and their response behavior can be understood. Our results suggest that there is no single rule capable of accounting for all individuals. Therefore, one goal of future models is to determine and use discriminative features enabling the detection of the reasoning strategy most fitting to a specific reasoner.

## Acknowledgements

This paper was supported by DFG grants RA 1934/3-1, RA 1934/2-1 and RA 1934/4-1 to MR.

## References


Baratgin, J., Douven, I., Evans, J. S. T., Oaksford, M., Over, D., & Politzer, G. (2015). The new paradigm and mental models. *Trends in Cognitive Sciences*, *19*(10), 547–548.

Chapman, L. J., & Chapman, J. P. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology*, *58*(3), 220–226.

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

Copeland, D. E. (2006). Theories of categorical reasoning and extended syllogisms. *Thinking & Reasoning*, *12*(4), 379–412.

Johnson-Laird, P. N. (1975). Models of deduction. *Reasoning: Representation and process in children and adults*, 7–54.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, *16*(1), 1–61.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*(4), 646–678.

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

Morley, N. J., Evans, J. S. B. T., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *The Quarterly Journal of Experimental Psychology Section A*, *57*(4), 666–692.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, *102*(3), 533–566.

Ragni, M., Dames, H., Brand, D., & Riesterer, N. (2019). When does a reasoner respond: Nothing follows? In A. Goel, C. Seifert, & C. Freska (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society.* Montreal: Cognitive Science Society.

Ragni, M., Khemlani, S., & Johnson-Laird, P. (2014). The evaluation of the consistency of quantified assertions. *Memory & Cognition*, *42*(1), 1–14.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, *120*(3), 561–588.

Revlin, R., Leirer, V., Yopp, H., & Yopp, R. (1980). The belief-bias effect in formal reasoning: The influence of knowledge on logic. *Memory & Cognition*, *8*(6), 584–592.

Revlis, R. (1975). Two models of syllogistic reasoning: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior*, *14*(2), 180–195.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Mit Press.

Wetherick, N. E., & Gilhooly, K. J. (1995). 'Atmosphere', matching, and logic in syllogistic reasoning. *Current Psychology*, *14*(3), 169–178.

Woodworth, R. S., & Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, *18*(4), 451–460.