



This article is part of the topic “Extending Rationality,” Emmanuel M. Pothos and Timothy J. Pleskac (Topic Editors).

Logic, Probability, and Pragmatics in Syllogistic Reasoning

Michael Henry Tessler,^{a,b} Joshua B. Tenenbaum,^a Noah D. Goodman^{b,c}

^a*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*

^b*Department of Psychology, Stanford University*

^c*Department of Computer Science, Stanford University*

Received 20 April 2021; received in revised form 11 November 2021; accepted 16 November 2021

Abstract

Syllogistic reasoning lies at the intriguing intersection of natural and formal reasoning of language and logic. Syllogisms comprise a formal system of reasoning yet make use of natural language quantifiers (e.g., *all*, *some*) and invite natural language conclusions. The conclusions people tend to draw from syllogisms, however, deviate substantially from the purely logical system. Are principles of natural language understanding to blame? We introduce a probabilistic pragmatic perspective on syllogistic reasoning: We decompose reasoning with natural language arguments into two subproblems: language comprehension and language production. We formalize models of these processes within the Rational Speech Act framework and explore the pressures that pragmatic reasoning places on the production of conclusions. We test our models on a recent, large data set of syllogistic reasoning and find that the selection process of conclusions from syllogisms are best modeled as a pragmatic speaker who has the goal of aligning the beliefs of a naive listener with those of their own. We compare our model to previously published models that implement two alternative theories—Mental Models and Probability Heuristics—finding that our model quantitatively predicts the full distributions of responses as well as or better than previous accounts, but with far fewer parameters. Our results suggest that human

Correspondence should be sent to Michael Henry Tessler, Department of Brain and Cognitive Sciences, Building 46, Room 3027, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA. E-mail: tessler@mit.edu

syllogistic reasoning may be best understood not as a poor approximation to ideal logical reasoning, but rather as rational probabilistic inference in support of natural communication.

Keywords: Pragmatics; Reasoning; Rational Speech Act; Syllogisms; Semantics

1. Introduction

Imagine discussing a recent tax proposal TR158 with a friend. Your friend remarks that in TR158:

Some of the taxes on the wealthy will be modified.

All of the tax modifications in the bill will expire in 10 years.

From these remarks, it follows—logically, necessarily—that *Some of the taxes on the wealthy will expire in 10 years*. That is, in all situations in which the premises are true, the conclusion is also true. The argument is logically valid. Now consider a superficially similar set of remarks: In TR158,

All of the taxes on the wealthy will be modified.

Some of the tax modifications in the bill will expire in 10 years.

You may be tempted to draw a similar conclusion: *Some of the taxes on the wealthy will expire in 10 years*. The argument seems reasonable, and yet the conclusion is not true in all situations in which the premises are true. Consider, for instance, that there could be tax modifications applied to the non-wealthy, and that those modifications are the ones that will expire in 10 years. Then, the premises would be true but the conclusion false: The argument is logically invalid.

Human intuitions about what conclusions reasonably follow from a set of premises seem to be supported by something other than pure logical necessity. Aristotle's invention of syllogisms—the kinds of two premise arguments shown above—can be interpreted, in fact, as an attempt to impose a simple rational structure on what might otherwise seem like fuzzy or messy human intuitions about logical relationships. In the modern era, cognitive psychologists have turned the syllogism back on the human mind, as a rational basis from which to evaluate human reasoning (Störring, 1908). Perhaps, unsurprisingly, how everyday human beings reason with syllogisms differs from what classical rules of logic would dictate (for a meta-analysis, see Khemlani & Johnson-Laird, 2012).

A syllogism is a formal logical argument that uses quantifiers (e.g., *all*, *some*) to relate two properties (e.g., *taxes on the wealthy* and *expiring in 10 years*) via a third ("middle") property (e.g., *being modified by the proposal*). Many theoretical frameworks have been brought to explain how humans reason with syllogisms. Logic or syntactic-based approaches have attempted to articulate modified logical rules—so-called "natural logic"—that humans deploy when evaluating syllogistic arguments (Braine & Rumain, 1983; Geurts,

2003; Rips, 1994). Process accounts have been developed to describe how we manipulate model- or diagram-based mental representations to arrive at a conclusion, the most prominent example being the Mental Models Theory (Johnson-Laird, 1975, 1983; Johnson-Laird, Khemlani, & Goodwin, 2015). Still, other approaches attempt to revisit the rational or functional basis of human logical reasoning, viewing deductive reasoning as a special case of probabilistic inference (Hahn & Oaksford, 2007; Oaksford & Chater, 2001, 2007; Tenenbaum, Griffiths, & Kemp, 2006); this approach is most prominently formalized in the Probability Heuristics Model of syllogistic reasoning (Chater & Oaksford, 1999).

Though the substantial prior work provides insight into how humans reason with syllogisms, a unified, explanatory theory of human syllogistic reasoning remains outstanding (Khemlani & Johnson-Laird, 2012). The most prominent accounts of syllogistic reasoning—Mental Models Theory (Johnson-Laird, 1975; Johnson-Laird et al., 2015) and the Probability Heuristics Model (Chater & Oaksford, 1999)—assume a standard competence-level theory involving first-order logic and probabilities, but, in order to explain human judgments, require the addition of either process-level idiosyncrasies or heuristics presumed to arise from resource constraints. Furthermore, the fact that syllogisms involve natural language demands a theory that can be integrated with principles of natural language—semantics and pragmatics—which are also likely central in human reasoning (Mercier & Sperber, 2017; Sperber & Wilson, 1986). Natural language semantics plays a role in many theories of syllogistic reasoning (Geurts, 2003; Johnson-Laird & Steedman, 1978; Khemlani & Johnson-Laird, 2012), but how those semantics interface with pragmatic reasoning has either been posited in an ad-hoc manner (Chater & Oaksford, 1999) or has been limited in scope (e.g., restricted to simple Gricean enrichment of the premises; Roberts, Newstead, & Griggs, 2001).

We take up the task of developing a competence-level theory that builds upon first-order logic and probabilities but construes syllogistic reasoning as part of communication. Our approach is at the level of a computational theory (in the sense of Marr, 1982) or rational analysis (in the sense of Anderson, 1990) that seeks to model the computations underlying human syllogistic reasoning without stipulating a precise implementation of such a reasoning process. In particular, we view syllogistic reasoning through the lens of the language understanding processes that guide everyday reasoning with language, a process that involves reasoning about how a listener would interpret a speaker's statements (Clark, 1996; Grice, 1975; Levinson, 2000). Specifically, we propose that human reasoning with syllogisms is composed of two key subproblems: A reasoner first interprets the premises of the syllogism by acting as a listener interpreting natural language expressions, and then the reasoner produces a conclusion by acting as a speaker deciding what to say. We formalize this hypothesis in a probabilistic model of pragmatic reasoning in the Rational Speech Act (RSA) framework (Frank & Goodman, 2012; Goodman & Frank, 2016; Scontras, Tessler, & Franke, 2018). Our model explains a number of qualitative phenomena in syllogistic reasoning as well as articulates a rational mechanism by which a person could respond *nothing follows*, an outstanding problem in theories of syllogistic reasoning (Riesterer, Brand, Dames, & Ragni, 2020). We additionally show that pragmatics enters into the conclusion selection process by

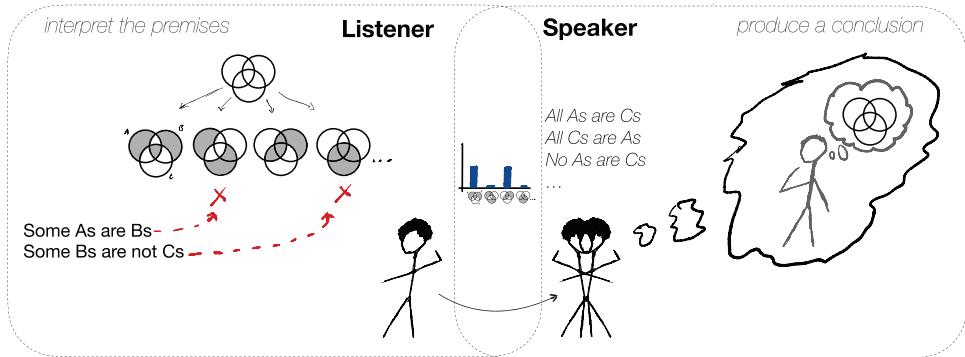


Fig. 1. Overview of the probabilistic pragmatics perspective of syllogistic reasoning. The reasoner first acts as a listener (left), interpreting the premises of the syllogism by updating their prior beliefs about the state (Venn diagram) according to the truth-functional meanings of the quantifier relations in the premises, resulting in a posterior distribution over states. Then, the reasoner acts as a speaker (right), deciding among a set of conclusions that would be the best to produce. The model developed in this paper restricts the listener model to only literal (though potentially noisy) interpretations of the premises, while the speaker model entertains pragmatic considerations (i.e., by reasoning about a listener).

way of participants reflecting on how well their choice of conclusion would convey their own beliefs to a naive listener. While there are almost certainly important processing details driven by resource limitations, we do not need to appeal to them to explain many aspects of human judgments. And though we do not explain all of the variance in the human data, the model provides a path forward for addressing several outstanding issues including how pragmatics might enter into the interpretation of premises and how the content of a syllogism can affect reasoning (known as *belief bias*; Dube, Rotello, & Heit, 2011; Klauer, Musch, & Naumer, 2000; Newstead, Pollard, Evans, & Allen, 1992; Oakhill & Garnham, 1993).

2. Rational Speech Act models

We develop a family of models of syllogistic reasoning as communication. A reasoner first acts as a listener who updates their prior beliefs about the world given the premises of the syllogism; then, the reasoner acts as a speaker who is deciding what conclusion to say given their updated beliefs about the world (Fig. 1). We describe our model in three stages: the ontology and semantics in the model, the listener model (or, how the premises get interpreted), and the speaker model (or, how the conclusion gets produced). In general, our approach is to have pragmatic inference (not only literal interpretation) be a component of both our listener and speaker models (cf., Roberts et al., 2001); in this paper, for simplicity, we only develop the pragmatic aspects of the probabilistic speaker model, restricting our probabilistic listener model to literal (though potentially noisy) interpretations.

2.1. Ontology and semantics

The relations used in classical Aristotelean syllogisms are quantifiers: *all*, *some*, *some ... are not*, and *none*. An example Aristotelean syllogism reads:

All artists are bakers.
 Some bakers are chemists.

Some artists are chemists.

The logical space defined by classical, Aristotelean syllogisms is composed of all combinations of quantifiers and term orders for the premises (e.g., QUANTIFIER A–B vs. QUANTIFIER B–A).¹ In total, each of the two premises can appear with one of the four quantifiers and two term orders yielding: 4 quantifiers \times 2 term orderings \times 2 premises = 64 syllogistic premises. Of the 64 unique syllogistic premises, only 27 yield a logically valid conclusion that uses the Aristotelean quantifiers (*all*, *some*, *some...not*, *none*): a conclusion that is true in every situation in which the premises are true.² For the remaining 37, several conclusions may follow from the premises, but there are no conclusions that are necessarily true given the premises; in terms of the syllogistic reasoning task, the “correct response” for these syllogistic premises is typically taken to be that “nothing follows” (or, “No Valid Conclusion” [NVC]).

Our model is built on top of an ontology of states (or, situations), which encode the types of objects that exist in a world. An object type is represented by its properties. Consider the second syllogism from above:

All artists are bakers.
 Some bakers are chemists.

Some artists are chemists.

One possible state that is consistent with this syllogism is:

```
artist baker
artist baker chemist
  baker chemist
```

In this state, there are three unique types of objects or entities: an artist who is also a baker, a baker who is also a chemist, and artist who is also a baker and chemist.³

This representation of states, in which only the unique object types are present, can also be displayed as a Venn diagram (Fig. 2). A region in the diagram corresponds to an object type (e.g., an A that is also a B and a C) and the Venn diagram contains eight regions $r \in R$, which correspond to all object types that could be referred to in the syllogism: $\{\{A, B, C\}, \{A, B, \neg C\}, \{A, \neg B, C\}, \{\neg A, B, C\}, \{A, \neg B, \neg C\}, \dots\}$. For simplicity, when possible, we will refer to regions by the properties that are present, omitting the absent properties; for example, $\{A, B, \neg C\}$ will be referred to as AB . Each region in the Venn diagram can take on a

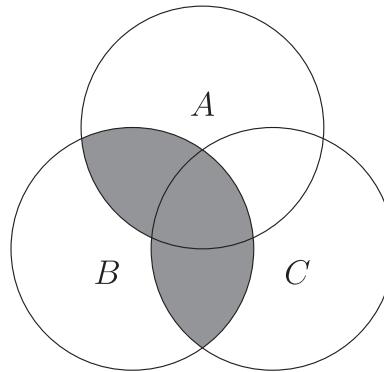


Fig. 2. A state is represented by Boolean values over the regions of a Venn diagram, corresponding to the presence or absence of unique object types. The diagram depicts a state in which there are ABs, BCs, ABCs, and no other object types.

Table 1

Logical meanings of example syllogistic sentences with examples of states that are consistent with the literal meanings and inconsistent with the literal meanings

Example Syllogistic Sentence	Literal Meaning	Consistent State	Inconsistent State
All As are Bs	$\forall i \in R : A(i) \Rightarrow B(i)$	{AB, ABC}	{A, ABC}
Some As are Bs	$\exists i \in R : A(i) \Rightarrow B(i)$	{A, AB}	{A, AC}
Some As are not Bs	$\exists i \in R : A(i) \not\Rightarrow B(i)$	{A, AB}	{AB, ABC}
No As are Bs	$\forall i \in R : A(i) \not\Rightarrow B(i)$	{A, AC}	{A, AB}

binary value, indicating the presence or absence of that unique object type. Then, a state or situation (Venn diagram) $s \in S$ is composed of the set of object types (regions) that are present (e.g., $\{AB, BC, ABC\}$ is one possible state). The set of unique states then has size $2^8 = 256$; however, the empty region (or, the object type that has neither of the three properties: $\{\neg A, \neg B, \neg C\}$) does not impact the truth conditions of the quantifier sentences used in syllogisms, and so we consider only $2^7 = 128$ states.

The classical syllogisms are composed of two premise arguments where each premise relates two properties via a quantifier. The quantifiers in classical syllogism are *all*, *some*, *none*, and *not all*.⁴ We use the classic logical semantics of these quantifiers described in Table 1. We assume that the semantics of *All As are Bs* describes a situation in which there exist objects that have the first property, sometimes referred to as the “existential presupposition” (e.g., *All As are Bs* is false if there are no As, and hence *All As are Bs* entails *Some As are Bs*).

2.2. Acting as a listener, interpreting the premises

We now introduce a probabilistic listener model that updates their beliefs about the states given the premises of the syllogism. The model’s prior on states $P(s)$ is constructed by

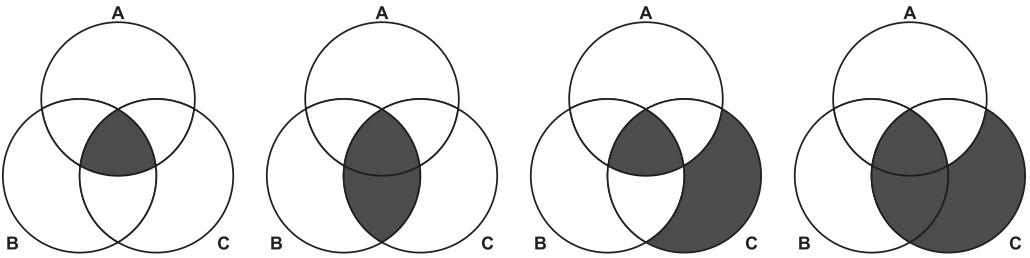


Fig. 3. Set of Venn diagrams (states) literally compatible with the premises of the logically valid syllogism: *All As are Bs, All Bs are Cs*. The conclusion relation *All As are Cs* is true in every compatible state.

sampling a Bernoulli random variable for each of the $r \in R$ regions: $P(s) \propto \prod_{r \in R} P(r)$. The premises of a syllogism update this prior on states via the literal, truth conditional meaning of each of the premises u (as shown in Table 1):

$$L_0(s | u) \propto P(s) \cdot \mathcal{L}(u, s) \quad (1)$$

where $\mathcal{L}(u, s)$ is the lexicon that encodes the literal meanings of the quantified utterances used in syllogisms. These meanings can resolve to deterministic outcomes (e.g., $\mathcal{L}(u, s) = 1$ if the utterance u is true of s , and 0 otherwise). In practice, however, we assume a small amount of noise ϕ in the semantics such that logically true $\mathcal{L}(u, s)$ may be judged false with probability proportional to $1 - \phi$; conversely, states that are literally incompatible with the utterance u have a small (ϕ) probability of being judged true (rather than a 0 probability). As a result, the semantics in the model take on continuous values such that $\mathcal{L}(u, s) = 1 - \phi$ or ϕ , a technique introduced by Degen, Hawkins, Graf, Kreiss, and Goodman (2020).

The model of literal interpretation can be used to update the prior distribution on states (Venn diagrams) into a posterior distribution given the meaning of the premises of the syllogism. For the interpretation of a syllogism, u is composed of two utterances (the two syllogistic premises): u_1 and u_2 . Belief updating proceeds by assuming both utterances are true:

$$L_0(s | u_1, u_2) \propto P(s) \cdot \mathcal{L}(u_1, s) \cdot \mathcal{L}(u_2, s) \quad (2)$$

Eq. 2 defines a probability distribution over states (Venn diagrams) given the premises of the syllogism. Logically valid syllogisms give rise to a distribution over states where a quantifier relationship between the conclusion terms of the syllogism is true in all possible states given the premises. For example, in the logically valid syllogism *All As are Bs, All Bs are Cs* (the “Barbara” syllogism), the relationship *All As are Cs* is true in every possible state in which the premises are true; in particular, the region ABC (an object that has all three properties) is true in every possible state (Fig. 3). On the other hand, the premises of logically invalid syllogisms tend to give rise to many possible states in which no particular syllogistic conclusion is true in every state; some conclusions are true in more states, however, than other conclusions, depending in general on the prior distribution over states $P(s)$. For example, in the logically invalid *All As are Bs, All Cs are Bs*, the conclusion *No As are Cs* is logically possible though it is true in fewer states than *Some As are Cs* (Fig. 4). The fact that our model

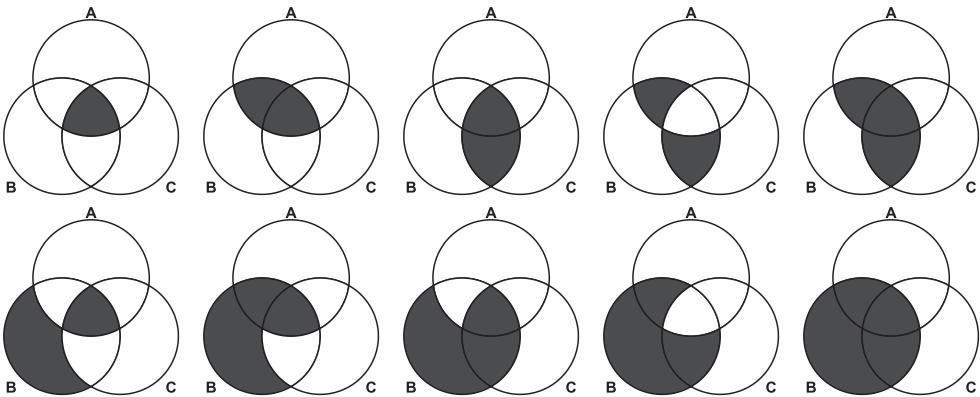


Fig. 4. Set of Venn diagrams (states) literally compatible with the premises of the logically invalid syllogism: *All As are Bs, All Cs are Bs*. No Aristotelean syllogistic conclusion is true in every compatible state.

reasons about a distribution of states that are likely given the premises enables it to draw conclusions from sets of premises that have no logically necessary (i.e., valid) conclusions.

2.3. Acting as a speaker, producing a conclusion

The second component of a syllogistic reasoning model is a model of conclusion production. That is, given a set of beliefs about the likely state of affairs (Eq. 2), a speaker decides what conclusion should be drawn. We formalize three models of conclusion production. The first is a literal speaker, who selects conclusions that are true given the speaker's beliefs about the state. The second is a pragmatic speaker, who selects a conclusion that would best convey information about an individual likely state to a naive listener (*State communication*). Our third and most sophisticated model is of a pragmatic speaker who chooses utterances to convey information about her entire belief distribution over states (*Belief alignment*).

2.3.1. Space of conclusions

The syllogistic reasoning task presents participants with a set of options for the quantifier relationship between the two terms of the syllogisms that are not described explicitly in relation to each other in the premises (i.e., the terms *A* and *C* if the premises are of the form *A-B, B-C*). Particular tasks differ in the set of options that are given to participants, but the maximal set includes the four quantifiers (*all, some, not all, none*) crossed with the two orderings of the terms of the conclusion (i.e., *A-C* or *C-A*). Thus, there are eight quantified conclusion options.

A ninth unique option, however, is also available: the option that *nothing follows* (sometimes described as *no valid conclusion*). The *nothing follows* conclusion is the logically correct answer for the logically invalid syllogisms and is produced to variable extents by participants in both logically valid and invalid syllogisms (Khemlani & Johnson-Laird, 2012). While the semantics of the quantifier conclusions are quite clear (Table 1), the proper

treatment of the *nothing follows* conclusion is not obvious. Extant theories treat this option indirectly, either as a by-product or last resort of a reasoning process, when it is treated at all (Ragni, Dames, Brand, & Riesterer, 2019; Riesterer et al., 2020). We approach the question of the meaning of *nothing follows* from a language production standpoint, where we formalize the statement that *nothing follows* as being tantamount to not saying anything at all. Formally, we model the *nothing follows* statement as a vacuous, or silent, utterance: an utterance that is true in every possible state of affairs. Hence, the impact of using such an utterance is that it does not serve to update a listener's beliefs about the state.

2.3.2. Literal production of conclusions: Literal Speaker

Given a belief updating model based on the premises of a syllogism L_0 , we define a literal production model S_0 that uses the listener L_0 's posterior distribution on states to determine what conclusion is likely to be true given the premises. This model S_0 can be viewed as one that is sampling a state from the listener's posterior distribution on states and randomly selecting among the conclusions that are literally true of that state. For a tighter comparison to the other models, we include a soft-max temperature parameter α that interpolates between probability matching behavior and probability maximizing behavior.

$$S_0(u_3 | u_1, u_2) \propto \exp[\alpha \cdot \sum_s \mathcal{L}(u_3, s) \cdot L_0(s | u_1, u_2)] \quad (3)$$

Here, $\mathcal{L}(u_3, s)$ is assumed to be a deterministic semantic function (unlike that inside L_0 , which is a function of noise parameter ϕ) that conditions on the literal meaning of the utterance u_3 .

2.3.3. Informative production of conclusions: State Communication

We define a more sophisticated speaker model who selects conclusions in order to convey information to a naive listener (who has heard neither premises nor conclusion). Since the reasoner does not know the state (but only has beliefs about the state given the premises of the syllogism), we operationalize the informational utility of a conclusion by marginalizing over the speaker's beliefs about the state. This speaker model first samples a state from the conditional distribution on states given the premises (as did the literal speaker in Eq. 3). Following standard practice in the RSA modeling framework, the utility of the utterance for a particular state is the negative surprisal of the state given the utterance. The speaker then chooses utterances soft-max optimally with respect to that utility function.

$$S_1(u_3 | u_1, u_2) \propto \exp [\alpha \cdot \sum_s \ln L_0(s | u_3) \cdot L_0(s | u_1, u_2)] \quad (4)$$

Note that the two L_0 's in Eq. 4 are actually two different agents. $L_0(s | u_1, u_2)$ is the reasoner who is acting as a listener updating their beliefs about the state given the premises. $L_0(s | u_3)$ is a hypothetical listener that the reasoner, when acting as a speaker, is communicating with; this hypothetical listener updates their beliefs about the state from the conclusion (u_3). The priors and likelihoods of these agents are equivalent and thus are represented equivalently.

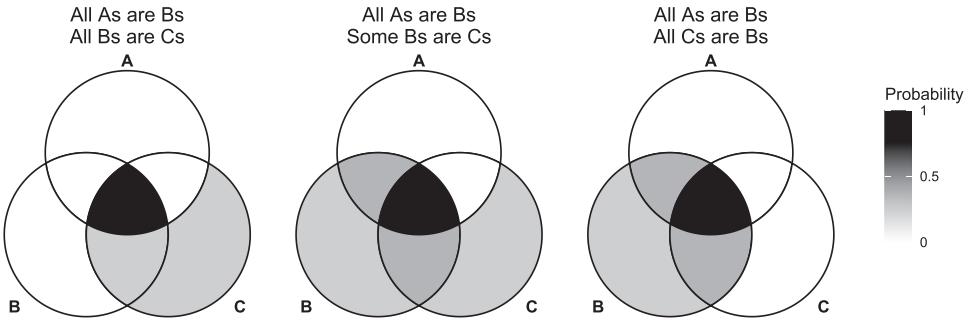


Fig. 5. A representation of the literal listener model's belief distribution over states. The diagram shows the marginal posterior means over regions. The opacity in each region denotes the posterior probability of that region (object type) being true of the premises. The probability of a particular state is proportional to the joint probability of each region being true.

2.3.4. Informative production of conclusions: Belief Alignment

Rather than the speaker sampling a state and then selecting an utterance to describe that state, the informational utility can be defined with respect to the reasoner's full belief distribution over states (Fig. 5; Goodman & Stuhlmüller, 2013; Scontras et al., 2018). The utility of a conclusion then is quantified by how well it would align the listener's beliefs with those of the reasoner (speaker). Formally, we use the information-theoretic measure of Kullback–Leibler (KL) divergence as the measure of alignment of the naive listener and reasoner's (speaker's) belief distributions. We denote the reasoner's belief distribution over the state as $S = L_0(s | u_1, u_2)$.

$$\text{KL}(S || L_0(u)) = \sum_s S(s) \log \frac{S(s)}{L_0(s | u)} \quad (5)$$

Then, the speaker model that selects conclusions in an informative manner is given by

$$S_1(u_3 | u_1, u_2) \propto \exp[\alpha \cdot -\text{KL}(L_0(s | u_1, u_2) || L_0(s | u_3))] \quad (6)$$

where $L_0(s | u)$ is the literal listener model defined in Eq. 1.

A critical ingredient to this belief alignment model of conclusion production is the noisy semantics we assume inside the literal listener L_0 . Without it, the speaker would be required to say *nothing follows* for any logically invalid syllogism. A logically invalid syllogism is one in which no conclusion is true in every state in which the premises are true. If a reasoner/speaker were to choose one of these conclusions, then the states that are incompatible with the conclusion would be imagined (by the hypothetical literal listener) to have 0 probability. But since the conclusion is not logically valid, the reasoner/speaker believes these states to have non-0 probability. This combination leads Eq. 5 to be undefined, which results in the model always saying *nothing follows* for logically invalid syllogisms. By adding a small amount of noise into the semantics, we avoid this boundary case.

3. Modeling the Ragni et al. (2019) data set

We test our models using a dataset published by Ragni et al. (2019). The data set consists of the results of a web experiment in which participants ($n = 139$) provided conclusions to all 64 syllogisms. In the experiment, participants completed all syllogistic reasoning problems after a brief training phase. Participants responded by selecting one of the nine possible responses for the conclusion of the syllogism (4 quantifier choices \times 2 term orders + 1 *nothing follows*). We use this data set because it is large and comprehensive (i.e., has many responses for all 64 syllogisms).

3.1. Bayesian data analysis

To assess how well our models can accommodate the syllogistic reasoning data from Ragni et al. (2019), we embed each model variant (Literal Speaker, State Communication, Belief Alignment) within a Bayesian data analysis model to infer the credible values of the latent parameters of the models and generate predictions given those parameters (Fig. 6).

3.1.1. Parameters and priors

The RSA modeling framework assumes that speakers are approximately rational agents who produce utterances as actions whose utility is defined by conveying information to a listener. The degree of rationality of the speaker interpolates between probability matching behavior and utility maximizing behavior. For the speaker rationality parameter α , we assume a uniform prior over the range of (0, 10), consistent with the prior literature on RSA models. We additionally assume a small amount of noise in the literal semantics of the utterances such that with a small probability ϕ , the literal listener disregards the utterance heard and does not update its beliefs via that utterance; we put a uniform prior over the limited range of (0, 0.1) over this parameter since we assume the noise to be a small quantity. For simplicity, we parametrize the prior distribution over states $P(x)$ in the reasoner models in a strictly uninformative manner: all regions (object types) have equal probability of being true, which is fixed to 0.5.⁵

Finally, we assume one global “order bias” parameter that modulates the pragmatic reasoner’s prior preferences for conclusions with a specific ordering of the terms (A–C vs. C–A) based on the ordering of the terms in the premises. These preferences are sometimes referred to as “figural effects,” tracing back to the “figure,” or term orderings, of the premises (Rips, 1994; Wetherick & Gilhooly, 1990). Our goal is not to provide an explanation for these effects, but rather posit an effect in line with the observations of Johnson-Laird and Bara (1984), Johnson-Laird and Steedman (1978), Geurts (2003) that participants prefer conclusions whose subject term matches the subject of one of the premises. Formally, the conclusion prior is parameterized by unnormalized probability weights, where a uniform prior would have a weight of 1 for all conclusions. We introduce a global preference parameter β for conclusions whose order of the terms (A–C vs. C–A) matches the order of the terms in the premises. This first-term preference works by checking whether either A or C appear in the subject position of one of the premises. If only one appears in the subject (e.g., the syllogistic

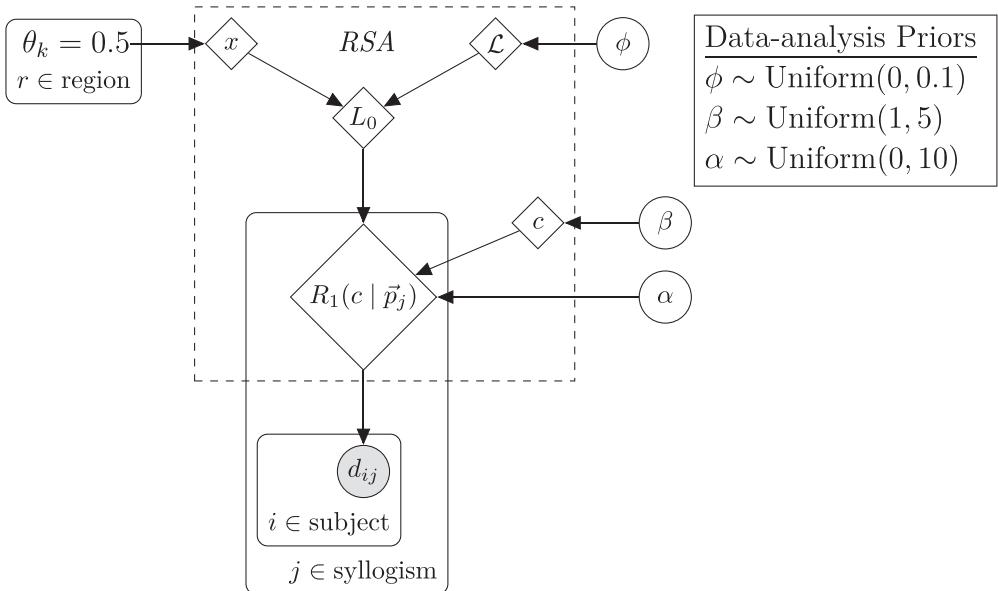


Fig. 6. Bayesian data analysis model for Rational Speech Act (RSA) model variants. The conclusion choices for each i participant on each j syllogism d_{ij} are generated via the pragmatic reasoner model R_1 , which defines a conditional distribution on conclusions c given premises \vec{p}_j . The reasoner model R_1 is an RSA model that interprets premises according to a literal listener model L_0 , which updates its beliefs according to the literal meaning \mathcal{L} of the sentences it hears. Literal meanings are the standard, truth-functional meanings of the quantifiers (Table 1) with a small amount of noise ϕ . The prior distribution over states $P(x)$ is parameterized by independent Bernoulli weights θ_k for each of the seven regions r of a Venn diagram, which we assume are all equal to 0.5. The reasoner R_1 's prior distribution over conclusions includes a preference β for term-orderings (A–C vs. C–A) that match the ordering of the terms in the premises. Finally, the reasoner R_1 selects conclusions according to the decision rule of the model variant (Literal Speaker, State Communication, Belief Alignment), each which has a soft-max parameter α .

frame A–B/B–C), then conclusions where that same term appears in the subject (e.g., A–C) will have a prior preference determined by β . If both or neither appear in the subject (e.g., A–B/C–B or B–A/B–C), then no preference is given in the conclusion prior, corresponding to an unnormalized probability weight of 1. We put a uniform prior between (1, 5) on this preference weight since we assume it to be a preference for producing conclusions that match the order of the terms of the syllogism.

3.1.2. Implementation and inference

We implemented the cognitive models and the Bayesian data analytic models in the probabilistic programming language WebPPL (Goodman & Stuhlmüller, 2014); model code and data analysis scripts can be found at <https://github.com/mhtess/syllogism-paper>. To learn about the credible values of the parameters, we ran three Markov chain Monte Carlo (MCMC) chains each for 15,000 iterations discarding the first 5,000 iterations for burn-in. We assessed convergence through visual comparison of the posterior distributions to ensure that the inferences that resulted from looking at different chains

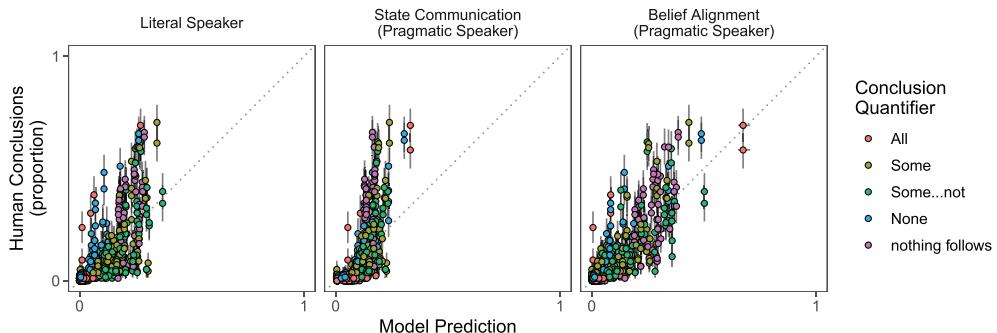


Fig. 7. Posterior predictive model fits for the literal speaker model and two versions of the pragmatic speaker model. Each point represents one of nine possible conclusions for one of the 64 possible syllogisms. Error bars denote bootstrapped 95% confidence intervals for the empirical data. Posterior model predictions did not exhibit significant variance; the horizontal error bars that are large enough to see represent the 95% Bayesian credible intervals for the model predictions.

were not appreciably different.⁶ We performed model comparison by computing Bayes Factors, which quantify the likelihood of the data under each model averaging over the model's prior distribution over parameters. By averaging over each model's prior on parameters, Bayes factors implicitly penalize models with extra degrees of freedom (Lee & Wagenmakers, 2014). We estimated the marginal likelihoods of the data under each model using an Annealed Importance Sampling algorithm (Neal, 2001).

3.1.3. Results

Overall, we found that the Belief Alignment pragmatic speaker model predicted the empirical data the best and provided the most parsimonious explanation of the three models (Fig. 7). This model performed several orders of magnitude better at explaining the data, taking into the account model complexity (in comparison to the Literal Speaker model, $\log BF \approx 300$; in comparison to the State Communication pragmatic speaker $\log BF \approx 1200$). It also achieved the highest correlation coefficient and lowest mean squared error of the RSA model variants (Table 2).

The Belief Alignment pragmatic speaker model captures a number of interesting patterns in the human syllogistic reasoning behavior. By incorporating a notion of informativeness into the selection of a conclusion, the model is able to show a preference among equally true conclusions. For example, in the *All As are Bs/All Bs are Cs* syllogism, the conclusion *Some As are Cs* is true in every possible world that *All As are Cs* is true, because *all* entails *some*. Nonetheless, the Belief Alignment pragmatic speaker model and human participants strongly prefer the *all* conclusion (Fig. 8, top-left facet); the model arrives at this preference because *all* more informatively conveys the belief distribution of the speaker in comparison to *some*. Syllogisms for which the premises provide little information about the conclusion also reveal interesting reasoning patterns in both participants and the models. In the *No As are Bs/No Bs are Cs* syllogism, participants tend to respond either *nothing follows* (the deductively correct conclusion) or either *No As are Cs* or *No Cs are As* (Fig. 9, second row, second column).

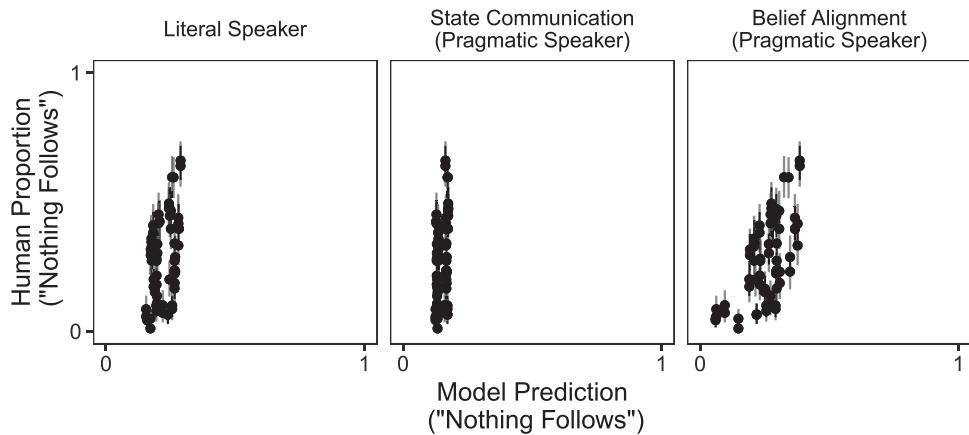


Fig. 8. Model predictions and behavioral results for the nine possible conclusions to 16 syllogisms. Syllogisms are of the form: As are Bs. Bs are Cs. Error bars denote bootstrapped 95% confidence intervals for the human data and 95% Bayesian credible intervals for the model predictions.

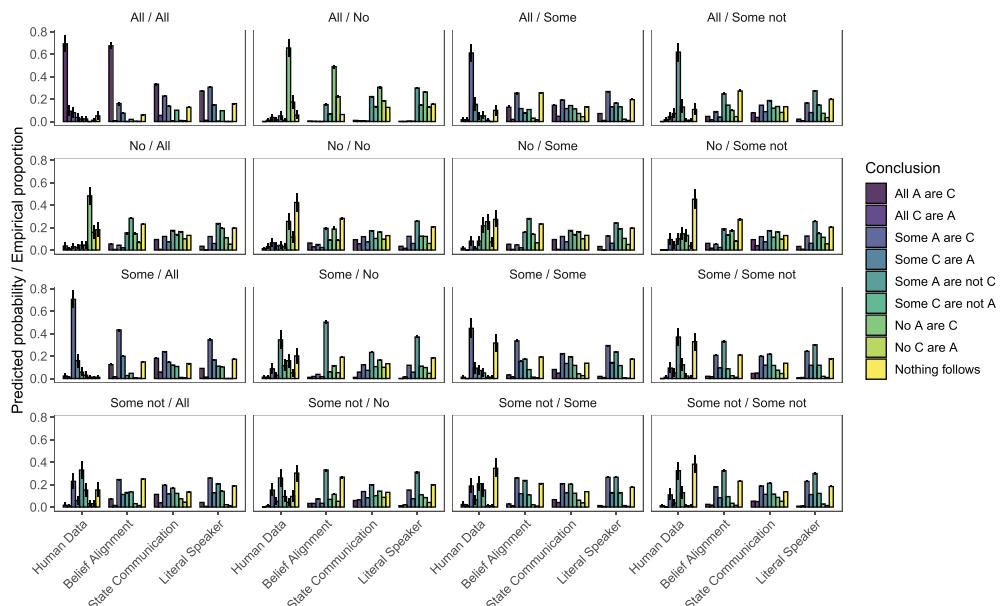


Fig. 9. Posterior predictive model fits for the *nothing follows* conclusion (a subset of the data shown in Fig. 7). The Belief Alignment model shows the greatest range of predictions, owing to the formalization of the *nothing follows* conclusion as a semantically-vacuous utterance. Error bars denote bootstrapped 95% confidence intervals for the empirical data. Posterior model predictions did not exhibit significant variance; the horizontal error bars that are large enough to see represent the 95% Bayesian credible intervals for the model predictions.

Table 2

Summary statistics for the RSA and alternative models predicting the data from Ragni et al. (2019). Alternative models were fitted to both group-level data and individual-level data. Individual-level variants use a new set of parameters for each individual participant in the data set ($n = 139$)

Model	Parameters	r	r^2	MSE
Belief Alignment	3	.82	.66	0.007
State Communication	3	.67	.45	0.013
Literal Speaker	3	.7	.48	0.011
mReasoner (group)	4	.71	.51	0.041
mReasoner (individual)	556	.82	.68	0.017
Probability Heuristics (group)	5	.77	.59	0.048
Probability Heuristics (individual)	695	.83	.69	0.02

The literal speaker model predicts what is likely to be true given the premises: *Some As are not Cs* is the most likely state of affairs. The State Communication model tries to be more informative, but because it only reasons about one particular state (Venn diagram) at a time, the more informative conclusion is to say *Some As are Cs*, at the same time downweighting the *nothing follows* conclusion. The Belief Alignment model takes a different approach: By trying to convey its entire distribution of belief states, the more informative conclusion it draws is either to say *No As are Cs* or *nothing follows*, most closely matching the human data.

The model is also able to adjust its conclusions based on the subtle shifts in the meaning of the premises. One interesting three-way comparison is the *Some As are Bs/Some Bs are Cs* vs. *Some As are Bs/Some Bs are not Cs* vs. *Some As are not Bs/Some Bs are not Cs* syllogism (Fig. 9, bottom-right quadrant). For *Some/Some*, the most likely and useful conclusion is *Some*. By changing the second premise to *Some Bs are not Cs* (*Some/Some not* syllogism), the modal conclusion changes to *Some not*. Yet changing the first premise also to *Some As are not Bs* (*Some not/Some not* syllogism) does not change the modal conclusion, which is also *Some not*. Both the Belief Alignment model and the Literal Speaker are able to capture these subtle changes, revealing that the literal meanings of the premises are shifting participants choice of conclusions.

By providing a substantive, information-theoretic hypothesis about the communicative function of the *nothing follows* conclusion, the Belief Alignment model exhibits the highest dynamic range among the RSA models in predictions about how and when people are likely to assert that an argument has no valid conclusion (Fig. 8). The model's predictions automatically accord with two intuitive boundary conditions: For syllogisms for which there is an informative and valid quantifier conclusion (e.g., *All As are Bs/All Bs are Cs*), the Belief Alignment model strongly disfavors the *nothing follows* response. For other syllogisms that have no logically valid conclusion and do not strongly imply other conclusions, the Belief Alignment model puts substantial probability mass on the *nothing follows* conclusion. Between these two extremes, and across the full data set of arguments, the belief alignment model explains a modest amount of the quantitative variability in people's preferences for the *nothing follows* conclusion ($r_{\text{belief}} = 0.64$, $r^2_{\text{belief}} = 0.40$, $\text{MSE}_{\text{belief}} = 0.017$) and more than

either the State Communication speaker model ($r_{state} = 0.31$, $r_{state}^2 = 0.10$, $MSE_{state} = 0.04$) or the Literal Speaker model ($r_{literal} = 0.44$, $r_{literal}^2 = 0.19$, $MSE_{literal} = 0.025$).

The Belief Alignment model's inferred parameter values provide more insight into how the model is able to make the predictions it does. The MAP and 95% credible interval for the figural effect (first-term preference) parameter was $\beta = 2.01(1.82, 2.14)$, showing that our model prefers conclusions whose terms match those of the premises by a factor of 2:1. The semantic noise parameter was inferred to be $\theta = 0.06(0.02, 0.06)$, showing that even a little bit of noise in the semantics aids the model in making better predictions. Finally, the speaker optimality parameter was $\alpha = 6.88(1.43, 7.15)$, covering a range of values consistent with what is typically observed in the RSA literature.

3.2. Comparison to other theories

Our modeling approach is a computational or rational level of analysis model (Anderson, 1990; Marr, 1982) that seeks to describe the computations that underly human syllogistic reasoning without stipulating a precise implementation of such a reasoning process. Many other theories of syllogistic reasoning instead consider the process by which a human reasoner solves a syllogism. The two most influential theories of human syllogistic reasoning are the Mental Models Theory (Johnson-Laird, 2006; Khemlani & Johnson-Laird, 2013) and the probability heuristics model (Chater & Oaksford, 1999), each of which share features with our logic, probability, and pragmatics approach.

3.2.1. Alternative models

Mental Models The Mental Models Theory considers reasoning over a set of iconic models, in which individual entities are represented explicitly, not unlike our state space of Venn diagrams. The reasoning process under this theory proceeds by first constructing a mental model, proposing a conclusion that can be drawn from such a model, and then engaging in a search for counterexamples to that conclusion; if a counterexample is found, either the mental model is revised and the search for a conclusion continues or the reasoner concludes *nothing follows*.

Here we compare to a recent implementation of this theory (the mReasoner model; Khemlani & Johnson-Laird, 2013), which is also probabilistic in nature (e.g., the model determines the number of objects to represent in the model by sampling from a Poisson distribution). The model has four free parameters (canonically denoted as λ , ϵ , σ , and ω): λ determines the number of entities represented in the model, ϵ determines the kind of mental model constructed (either a canonical or non-canonical representation of the entities consistent with the syllogism; as described by Bucciarelli & Johnson-Laird, 1999), σ denotes the probability of engaging in a search for counterexamples, and ω modulates the conclusion selection process that results when engaging in a search for counterexamples.

Probability Heuristics The Probability Heuristics model (PHM) of Chater and Oaksford (1999) seeks to account for human syllogistic reasoning as a process of

determining probabilistically valid (“p-valid”) conclusions, which can be approximated through a number of heuristics. Two heuristics generate candidate quantifiers for the conclusion, which use a notion of informativity for which Chater and Oaksford (1999) define a rank ordering: *all > some > none > some...not*. The “min-heuristic” takes the quantifier of the least informative premise to be a potential conclusion quantifier, and the “p-entailment heuristic” further allows for a quantifier that is entailed by the “min-heuristic” quantifier to also be a candidate quantifier for the conclusion (e.g., if *all* is generated from the “min-heuristic”, *some* will be generated from the “p-entailment heuristic”). A parameter (canonically referred to as p_{ent}) modulates which heuristic (min or p-entail) the reasoner uses to generate conclusions. A third heuristic (“attachment”) determines the ordering of the terms in the conclusion (A–C vs. C–A): this heuristic stipulates that the subject term of the conclusion should match the subject term of the least-informative premise, when allowable, similar in spirit to the β “order bias” parameter used in the Bayesian data analysis of the RSA models.

Two more heuristics determine the confidence (or, probability) of each of the candidate conclusions. Foremost, the “max-heuristic” stipulates the probability of a candidate conclusion is proportional to the informativity of the most informative premise. The “O-heuristic” additionally downweights conclusions that use the “some...not” quantifier. In the implementation of the Probability Heuristics Model that we use, these heuristics are parameterized by four parameters, which govern the confidence placed in conclusions of each of the four quantifiers (all, some, some...not, and none).

3.2.2. Quantitative comparison

To compare the quantitative accuracy of these alternatives theories to our own, we analyzed the results of a recent study of these theories that focused on modeling the behavior of individual participants in the syllogistic reasoning task (Riesterer, Brand, & Ragni, 2020). This study explored the predictive capabilities of the mReasoner and probability heuristics theories by treating the parameters of the models as individual-level parameters that could vary among participants (as well as the group-level version of the model, in which a single set of parameters governs the behavior of the population of participants). Riesterer et al. (2020) fit the parameters of these models to the same data set from Ragni et al. (2019) that we modeled above with the RSA models.⁷

Since Riesterer et al. (2020) focus on predictive accuracy of the models for individual behavior, their implementation of the Probability Heuristics Model treats the five model parameters as binary (0 or 1). This focus on predictive accuracy means that if a participant uses a heuristic < 50% of the time, the best predictions that could be made would result by setting the corresponding parameter to 0. To examine the predictions of these theories in a manner more comparable to our analysis of the RSA models, we treated the predictions of these alternative models for individual participants similar to how we treated the behavioral responses of individual participants. We constructed bootstrapped 95% confidence intervals for the model predictions by resampling with replacement the predictions for each unique syllogism and conclusion.

The mReasoner and Probability Heuristics models with parameters fit globally to the whole data set (group-level models) perform quite poorly overall (Fig. 10). The models largely make

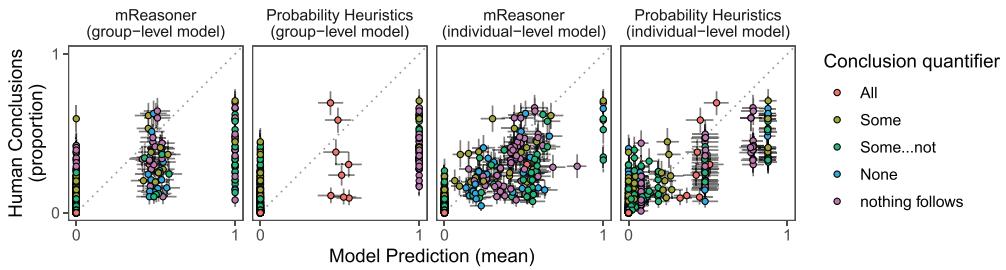


Fig. 10. Performance of alternative theories. Each point represents an individual premise and conclusion pair. Group-level models use parameters that are fit to the group-level data. Individual-level models use parameters specific to individual participants.

categorical predictions (1s or 0s), whereas the aggregated human data are graded. With parameters fit to individual participants and predictions aggregated, the mReasoner and Probability Heuristics models perform quite a bit better and comparable to our RSA models in terms of their correlation and mean squared errors, though the RSA models all perform better in terms of absolute mean squared error (Table 2). In addition, the individual-level parameterization of the mReasoner and Probability Heuristics models use a different set of parameters for each of the 139 participants in the task, yielding a total number of parameters for the model that is one-to-two orders of magnitude larger than the group-level versions of these models and our RSA models (Fig. 10).⁸ These results suggest that these alternative theories encoding process-level hypotheses may be well suited to model an individual participant's reasoning behavior, but are insufficient to adequately describe the computational problem that human reasoning is aiming to achieve.

4. General discussion

Reasoning with syllogisms sits at the intriguing intersection of language and logic. Syllogisms are a logical system, but they are expressed using natural language quantifiers and invite natural language inferences. We introduced a probabilistic pragmatic perspective for syllogistic reasoning, where reasoning involves behaving both as a listener interpreting the premises of the syllogism and as a speaker deciding what conclusion to say. We developed and tested a family of probabilistic models of syllogistic reasoning that incorporate a logical semantics with a generative model of situations and formalize how considerations of pragmatic informativeness could affect the selection of conclusions. We found that conclusions selected by participants were best explained by a model that chose its conclusion from a syllogism by reasoning about how well that conclusion would align a naive listener's beliefs with those of their own.

Our proposal is that human syllogistic reasoning is part logical, part probabilistic, and part pragmatic. Our theory is a rational model that articulates the functional problem human syllogistic reasoning is aiming to solve, which is at its core a problem of communication. Our work

builds on the goal articulated by Chater and Oaksford (1999) of building a rational model of syllogistic reasoning, and we do so without positing any heuristic solutions to the inference problems. Our model is also deeply connected to the Mental Models approach (Johnson-Laird, 1975; Johnson-Laird et al., 2015) in that the states of affairs that our model reasons over are mental models of the world. One subtle difference is that a mental models theory would typically reason about tokens or instances of categories, whereas our model reasons about types.

We contrasted two mechanisms by which pragmatic considerations can enter into the production of conclusions: one based on informatively describing a particular state (or, Venn diagram) and one based on conveying the reasoner's entire belief distribution over states (i.e., their distribution over possible Venn diagrams). The results, both quantitatively and qualitatively, provide overwhelming evidence that it is the latter notion of belief alignment that most closely tracks human reasoning patterns. This formalization of pragmatic considerations differs sharply from previous perspectives on pragmatics in syllogistic reasoning. Most notably, Roberts et al. (2001) explored the predictions of an approach most similar to our State Communication model: how a reasoner would choose Gricean-enriched quantifiers (e.g., *some* which implies not all) for each of the possible states (Venn diagrams) implied by the premises (which could have also been pragmatically enriched). Consistent with the observations of Roberts et al. (2001), such a State Communication model fails to provide any quantitative improvement across the data set, while our Belief Alignment model does.

A major shortcoming of extant approaches to syllogistic reasoning is that they do not provide a clear and interpretable mechanism by which the conclusion that *nothing follows* can be rationally produced (Riesterer et al., 2020). We formalized *nothing follows* as a semantically vacuous statement that does nothing to update the listener's beliefs. This kind of utterance is, in the semantic sense, always true, and thus, the literal speaker produces this utterance with constant probability.⁹ The Belief Alignment speaker, by contrast, selects utterances in order to convey their full beliefs about the state to the listener. When their beliefs are not substantially changed by hearing the premises (i.e., when the premises are uninformative relative to the reasoner's prior beliefs), the speaker will have a positive incentive to say *nothing follows* as that utterance would correspondingly not alter the hypothetical listener's beliefs, leading to strong alignment between speaker and listener beliefs. To our knowledge, this formalization is the first substantive hypothesis about how the *nothing follows* conclusion can be rationally produced in the syllogistic reasoning task.

4.1. Extending the framework

Our approach to syllogistic reasoning is at its core a Bayesian approach, where the premises of a syllogism update a reasoner's prior beliefs about the relationship between three properties. In our modeling, we did not assume anything substantive about these prior beliefs, except that they were constant across all syllogisms. Prior beliefs can, however, form one basis for formalizing hypotheses about how the content of the syllogism (i.e., the content of the particular terms: A, B, C) affects reasoning, a topic known as "belief bias" that has been of interest to psychologists for decades (Cherubini, Garnham, Oakhill, & Morley, 1998; Oakhill

& Garnham, 1993; Oakhill, Johnson-Laird, & Garnham, 1989). For example, Tessler (2015) found that some content effects in syllogistic reasoning could be captured by empirically measuring the relevant prior beliefs and using them in a model similar to that presented here. An alternative way in which a reasoner's beliefs could influence their reasoning is by a reasoner explicitly rejecting a premise (e.g., not accepting that *All artists are bakers*). A basic "accept/reject premise" mechanism is already included in our model via the semantic noise parameter ϕ , the probability with which the reasoner does not update their beliefs using a premise of a syllogism, which we also assumed to be a constant term.¹⁰ Even with a constant semantic noise parameter value (i.e., a constant noise in the likelihood), effectively different rejection rates for premises can arise via different prior beliefs about properties, due to Bayes' rule (e.g., the degree to which *All artists are bakers* is rejected depends on the prior beliefs about the properties of being an artist and being a baker).¹¹ Still, it is not clear that differential prior beliefs would yield all the flexibility that varying noise rates could generate. Further work should investigate these prior beliefs and differential noise rates to formalize hypotheses about how world knowledge rationally affects reasoning, which could speak to outstanding debates about whether these effects should be considered a "bias" (Dube et al., 2011; Garnham & Oakhill, 1986; Klauer et al., 2000).

Our model uses the literal, natural language semantics of the premises of the syllogism to update the model's prior beliefs. To extend our approach to reasoning with other explicit quantifiers (e.g., *most* and *few*), one could either generalize the binary regions in the Venn diagram to have continuous values or articulate an ontology of states that makes use of token entities instead of types (see, e.g., Tessler & Goodman, 2014, cf., Geurts, 2003). Additionally, syllogisms that appear in natural discourse often make use of implicitly quantified, generic statements (e.g., *Taxes on the wealthy will be modified*; see also Khemlani, Leslie, & Glucksberg, 2008); our approach could naturally be expanded to incorporate recent advances in the computational semantics of generics to account for syllogistic reasoning about generic premises (Tessler & Goodman, 2019).

Finally, this model could be used to consider the implications of pragmatic reasoning on the interpretation of the premises. Such a pragmatic interpretation model could be constructed as a standard RSA pragmatic listener model (Frank & Goodman, 2012; Goodman & Frank, 2016; Scontras et al., 2018); that is, treating the listener component of the model as a listener who reasons about why a speaker produced the premises that they did, where the speaker decides what utterance to produce to convey information about the state (or, belief distribution) to a naive literal listener. Indeed, we have explored several versions of this kind of pragmatic model in the course of this project (described briefly in the Appendix) but found that no model extension we tried provided a substantially better quantitative fit to Ragni et al. (2019) data than the model that interprets the premises literally, as we presented.¹² We do not take this null result as evidence that pragmatics does not enter into the interpretation process of a syllogism. We have only considered one particular experimental protocol and data set (from Ragni et al., 2019), where participants were presented with all 64 syllogisms; this experimental protocol potentially reveals to the participant the generative process of the syllogism (parametric recombination of quantifiers and term orders), which could induce a more logical analysis of the syllogisms. As well, the participants were from Western, edu-

cated, industrialized, rich, democratic societies, a feature not shared with most other human beings (Henrich, Heine, & Norenzayan, 2010). Reasoning with syllogisms varies appreciably across individuals (Bacon, Handley, & Newstead, 2003; Galotti, Baron, & Sabini, 1986) and substantially so in illiterate persons (Dias, Roazzi, & Harris, 2005). In other contexts or populations, we may see that reasoning with syllogisms as a special case of reasoning about an argument or a riddle (cf. Luria, 1976; Ong, 1982). Extending our modeling approach with a pragmatic interpretation component will be important for analyzing such behavior.¹³

4.2. Relationship to other frameworks for extending rationality

Human reasoning with syllogisms has captured the attention of theoretical and empirical psychologists from the time of Störring (1908), and many theoretical frameworks have been brought to bear on these capacities for the last half-century. Our approach extends the structured, Bayesian perspective on cognition (Goodman, Tenenbaum, & Gerstenberg, 2015; Tenenbaum, Kemp, Griffiths, & Goodman, 2011) with principles of natural language, formalized in models of rational communication (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013; Scontras et al., 2018). Many contemporary cognitive modeling frameworks are instead concerned with extending notions of rationality by incorporating computational constraints that the mind may operate within, and such approaches to these constraints might help explain the residual variability that our pragmatics models left unexplained. We briefly discuss the relationship of our approach to three approaches to extending rationality: resource rationality, quantum models of cognition, and ecologically rational heuristics.

A recent powerful framework for the integration of processing constraints into rational models of cognition—*resource rational analysis* (Griffiths, Lieder, & Goodman, 2015; Lieder & Griffiths, 2020)—could be brought to bear directly on our probabilistic pragmatics perspective in a number of ways. In contrast to our Belief Alignment speaker model who aims to communicate their entire belief distribution over states, a resource rational speaker might instead draw a few samples from its posterior distribution on states and seek to convey that collection of states succinctly (cf., Vul, Goodman, Griffiths, & Tenenbaum, 2014). Furthermore, a resource rational pragmatic reasoner (e.g., behaving as a pragmatic listener) may not fully represent the knowledge state or goals of their interlocutor (the hypothetical speaker who produced the syllogism) due to computational constraints, leading to asymmetries in how the listener/reasoner responds to the speaker’s argument versus how the speaker might have intended (Hawkins, Gweon, & Goodman, 2021); this kind of wrinkle introduced by resource rational considerations could have important implications for how the probabilistic pragmatic perspective bears on the problem of syllogistic reasoning.

Quantum probability models of cognition are similar to Bayesian models but take as their starting point the axioms of quantum probability theory (Pothos & Busemeyer, 2013). One key difference between quantum and classical theories of probability is the order dependence of conjunctive events under quantum theory. This lens, in particular, could be used for investigating deviations from Bayesian rational reasoning between logically identical syllogisms, whose premises are order reversed (e.g., *Some As are Bs/All Bs are Cs* vs. *All Bs are Cs/Some As are Bs*). At the same time, such divergences could also result from a probabilistic

pragmatics perspective, in which a speaker deciding to say two utterances in a particular order could lead to a non-independence of the likelihood of those utterances (see also footnote 11).

Heuristic accounts have also had success in explaining human syllogistic reasoning (e.g., the Probability Heuristics Model; Chater & Oaksford, 1999); such a perspective could be extended with the contemporary adaptive toolbox approach, by incorporating the interaction between the cognitive agent and the environment (Hertwig, Pleskac, & Pachur, 2019). Such an extension could be productive in describing the ecological setting in which human syllogistic reasoning can be best understood and how that might vary across culture. The heuristic perspective may, however, turn out to look a lot like a resource rational implementation of the Bayesian model (i.e., the line between an algorithmic implementation of rational inference and heuristics for inference may turn out to be superficial; see also, e.g., Lieder, Griffiths, Huys, & Goodman, 2018).

4.3. Conclusion

We articulate the verbal reasoning machinery that underlies the human capacity to reason with deductive arguments. This machinery is composed of three ingredients: logic, probability, and pragmatics. The combination of logic and probability enables us to reason about situations we have not experienced, but could encounter in the future. Pragmatics moves us outside of just what is plausible. Pragmatics helps us see the agent at the other end of the line, who hears our argument and decides for herself if the conclusion makes sense.

Acknowledgments

We thank Erin Bennett, Sophie Bridgers, Justine Kao, Max Kleiman-Weiner, Robert Hawkins, Daniel Hawthorne, and Long Ouyang for helpful conversations and insightful comment. We are also indebted to James T. Cargile of the University of Virginia for his incisive examination of the logical commitments of our project. This work was supported in part by NSF Graduate Research Fellowship DGE-114747 and NSF SBE Postdoctoral Research Fellowship Grant No. 1911790 to MHT, Army Research Office MURI Grant No. W911NF-19-1-0057 to JBT, and a Sloan Research Fellowship, ONR grant N00014-13-1-0788, and DARPA grant FA8750-14-2-0009 to NDG. The funders had no role in theory development, study design, modeling work, decision to publish, or preparation of the manuscript.

Notes

- 1 In addition to being restricted to considering only the Aristotelean quantifiers, the classic syllogistic reasoning task further only makes use of the primitive terms As, Bs, and Cs (e.g., one may not generate conclusions referring to NON-As).
- 2 Following Khemlani and Johnson-Laird (2012) and others, we describe the syllogisms assuming *existential import*; namely, that the premises describe categories that contain actual members (i.e., there are at least some artists, some bakers, and some chemists).

- 3 This formalism is different from a *mental model* in the style of Johnson-Laird (1983). In that formalism, one could have a model such as:

```

artist baker
artist baker chemist
    baker chemist
artist baker

```

where each row of the model denotes an individual entity that has the properties listed (e.g., the first row denotes an artist who is also a baker). In this model, there are two artist-bakers, one baker-chemist, and one artist-baker-chemist. The classical categorical syllogisms we model below involve the quantifiers *all* and *some* (and their logical negations); the truth conditions of these quantifiers (i.e., whether or not quantifiers are true) are not affected by the number of entities that have a particular combination of properties (i.e., it is inconsequential that there are two artist-bakers in the model above for the truth of the quantifiers *all*, *some*, *none*, or *some... not*). Hence, we can represent the model in a reduced format, in which only the unique object types are represented. For generalized quantifiers (e.g., *most*, *few*), the number of objects that have unique combinations of properties will matter. We discuss how our approach can be extended to reason with generalized quantifiers in the General Discussion.

- 4 These quantifiers are typically presented in sentences such that *none* is rendered as *no* (e.g., No As are Bs) and *not all* is rendered as *some __ are not* (e.g., some As are not Bs).
- 5 We experimented with more flexible parameterizations of the prior but found that these did not significantly improve model performance on this data set.
- 6 The full Bayesian data analytic model is unidentifiable owing to symmetries in the parameter space. Thus, a convergence statistic such as the Gelman–Rubin convergence statistic R (Gelman & Rubin, 1992) is not a good indicator for the reliability of the posterior predictions from these models. Accordingly, the R values for these MCMC chains were relatively high, indicating high chain variability. Notably, the R values for the posterior predictions were substantially lower than those for the parameters (most R values for the predictions fell between 1.3 and 1.4, whereas the R values for the parameters ranged between 1.75 and 2.5). The conclusions we draw thus are primarily about the model predictions and model comparison; less important (and less reliable) are the precise parameter estimates from these models.
- 7 The predictions of these models were accessed via the companion repository to the paper: <https://github.com/nriesterer/cogsci-individualization>.
- 8 We discuss model complexity here in a coarse-grained manner. The PHM parameters in the Riesterer et al. (2020) implementation are binary and thus do not add to model complexity in the same way as continuous parameters do. Furthermore, a hierarchical model for the individual subject-level parameters (i.e., partial pooling) could more thoroughly constrain the parameter space for the individual-level PHM model. Still, we think that the approximate complexity of the models can be appreciated by their order-of-magnitude difference in the number of parameters.

- 9 Technically, the Literal Speaker produces this vacuous utterance with probability inversely proportional to the number of conclusions that are semantically compatible with the premises, which will vary based on the logical validity of the syllogism.
- 10 The semantic noise parameter ϕ is a catch-all for many forms of uncertainty that the listener may hold, including but not limited to an explicit “accept/reject premise” inference, a listener’s uncertainty about the precise meaning of a premise, or other forms of uncertainty on the part of the listener.
- 11 With the noise probability ϕ , the literal listener does not update their beliefs and falls back on the prior, which could result in the states that are true of the premises receiving lower probability depending on whether they are *a priori* plausible or implausible. For example, given two correlated properties (e.g., *is a politician P, is corrupt C*) and constant noise in the semantics (ϕ), when a listener reads an unlikely premise (e.g., *No politician is corrupt*), the states with high posterior probability will be:
 - with probability $1 - \phi$ (the premise is true): $P, \neg C; \neg P, C; \neg P, \neg C$,
 - with probability ϕ (the premise is false): P, C (the state which has high probability under the prior).

If instead, the second property C were “selfless” (rather than “corrupt”), then the prior would provide more probability mass to the states other than P, C (since the two properties are relatively unlikely to co-occur). (This example should generalize to other quantifiers due to their logical symmetries: some = \neg none, some are not = \neg all.) The end result then is that states that are consistent with the premises would have differential posterior probabilities depending on if they are *a priori* plausible versus implausible, which is the same behavior in the model one would observe if the listener is explicitly rejecting versus accepting premises via differential noise rates.

- 12 Specifically, we formalized pragmatic listener models that interpret the premises from a rational speaker who wants to convey information about the full state as well as a speaker who wants to convey information about a number of different questions under discussion (QUD) (Roberts, 2004), in which the speaker only cares about conveying information about the A, C terms of the state; see also Tessler & Goodman (2014) for yet a different formulation of pragmatic interpretation of premises that achieves a comparable fit to that presented here.
- 13 It is additionally noteworthy that the pragmatic interpretation models we explored break the independence of the likelihood of utterances that is assumed by the literal interpretation model (Eq. 1). We believe that pragmatics provides a plausible mechanism by which the likelihood of otherwise independent utterances can become dependent (namely, via a rational speaker choosing to produce the two utterances).

References

- Anderson, J. R. (1990). *The adaptive character of thought*. New York: Psychology Press.
 Bacon, A., Handley, S., & Newstead, S. (2003). Individual differences in strategies for syllogistic reasoning. *Thinking & Reasoning*, 9(2), 133–168.

- Braine, M. D., & Rumain, B. (1983). Logical reasoning. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology*. Hoboken, NJ: Wiley.
- Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science*, 23(3), 247–303.
- Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, 258, 191–258.
- Cherubini, P., Garnham, A., Oakhill, J., & Morley, E. (1998). Can any ostrich fly?: Some new data on belief bias in syllogistic reasoning. *Cognition*, 69(2), 179–218.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A Bayesian approach to “overinformative” referring expressions. *Psychological Review*, 127(4), 591–621.
- Dias, M., Roazzi, A., & Harris, P. L. (2005). Reasoning from unfamiliar premises: A study with unschooled adults. *Psychological Science*, 16(7), 550–554.
- Dube, C., Rotello, C. M., & Heit, E. (2011). The belief bias effect is aptly named: A reply to Klauer and Kellen (2011). *Psychological Review*, 118(1), 155–163.
- Frank, M. C., & Goodman, N. D. (2012). Quantifying pragmatic inference in language games. *Science*, 336, 1–9.
- Galotti, K. M., Baron, J., & Sabini, J. P. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models? *Journal of Experimental Psychology: General*, 115(1), 16.
- Garnham, A., & Oakhill, J. V. (2005, April). Accounting for belief bias in a mental model framework: Comment on Klauer, Musch, and Naumer (2000). *Psychological Review*, 112(2), 519–520. <https://doi.org/10.1037/0033-295X.112.2.509>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472.
- Geurts, B. (2003). Reasoning with quantifiers. *Cognition*, 86(3), 223–251.
- Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11), 818–829.
- Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, 5(1), 173–84.
- Goodman, N. D., & Stuhlmüller, A. (2014). The design and implementation of probabilistic programming languages. <http://dippl.org>.
- Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2015). Concepts in a probabilistic language of thought. Margolis Eric, Laurence Stephen, *The conceptual mind: New directions in the study of concepts*. Cambridge, MA: MIT Press.
- Grice, H. P. (1975). Logic and conversation. Cole P., Morgan J. L., *Syntax and Semantics*. 3 New York: Academic Press, 41–58.
- Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, 7(2), 217–229.
- Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, 114(3), 704–732.
- Hawkins, R. D., Gweon, H., & Goodman, N. D. (2021). The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cognitive Science*, 45(3), e12926.
- Hawkins, R. X., Stuhlmüller, A., Degen, J., & Goodman, N. D. (2015). Why do you ask? Good questions provoke informative answers. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not weird. *Nature*, 466(7302), 29–29.
- Hertwig, R., Pleskac, T. J., & Pachur, T. (2019). *Taming uncertainty*. Cambridge: MIT Press.
- Johnson-Laird, P. N. (1975). Models of deduction. In R. J. Falmagne (Ed.) *Reasoning: Representation and process in children and adults* (pp. 7–54). Erlbaum.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Johnson-Laird, P. N. (2006). *How we reason*. Oxford: Oxford University Press.

- Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition*, 16(1), 1–61.
- Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, 19(4), 201–214.
- Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology*, 10, 64–99.
- Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, 111(33), 12002–12007.
- Khemlani, S., & Johnson-Laird, P. N. (2012). Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, 138(3), 427–57.
- Khemlani, S., & Johnson-Laird, P. N. (2013). The processes of inference. *Argument & Computation*, 4(1), 4–20.
- Khemlani, S., Leslie, S.-J., & Glucksberg, S. (2008). Syllogistic reasoning with generic premises: The generic overgeneralization effect. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, Vol. 30.
- Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, 107(4), 852–884.
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.
- Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43, 1–85.
- Lieder, F., Griffiths, T. L., Huys, Q. J., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review*, 25(1), 322–349.
- Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations*. Cambridge, MA: Harvard University Press.
- Marr, D. (1982). *Vision: The philosophy and the approach*. San Francisco: Freeman.
- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.
- Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing*, 11(2), 125–139.
- Newstead, S. E., Pollard, P., Evans, J. S., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition*, 45(3), 257–284.
- Oakhill, J., & Garnham, A. (1993). On theories of belief bias in syllogistic reasoning. *Cognition*, 46, 87–92.
- Oakhill, J., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, 31, 117–140.
- Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, 5(8), 349–357.
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.
- Ong, W. J. (1982). *Orality and literacy: The technologizing of the word*. London: Methuen and Company Ltd.
- Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, 36(3), 255–274.
- Ragni, M., Dames, H., Brand, D., & Riesterer, N. (2019). When does a reasoner respond: Nothing follows? In *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 2640–2546).
- Riesterer, N., Brand, D., Dames, H., & Ragni, M. (2020). Modeling human syllogistic reasoning: The role of “no valid conclusion”. *Topics in Cognitive Science*, 12(1), 446–459.
- Riesterer, N., Brand, D., & Ragni, M. (2020). Do models capture individuals? evaluating parameterized models for syllogistic reasoning. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*.
- Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: MIT Press.
- Roberts, C. (2004). Information structure in discourse. *Semantics and Pragmatics*, (5), 1–69.
- Roberts, M. J., Newstead, S. E., & Griggs, R. A. (2001). Quantifier interpretation and syllogistic reasoning. *Thinking & Reasoning*, 7(2), 173–204.

- Scontras, G., Tessler, M. H., & Franke, M. (2018). Probabilistic language understanding: An introduction to the rational speech act framework. <https://www.problang.org>.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*, Vol. 142. Cambridge, MA: Harvard University Press.
- Störring, G. (1908). In *Experimentelle untersuchungen über einfache schlussprozesse*. Leipzig: W. Engelmann.
- Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, 10(7), 309–318.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022), 1279–1285.
- Tessler, M. H.** (2015). Understanding belief bias by measuring prior beliefs for a bayesian model of syllogistic reasoning. In *Proceedings of ESSLLI* (pp. 225–237).
- Tessler, M. H., & Goodman, N. (2014). Some arguments are probably valid: Syllogistic reasoning as communication. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, Vol. 36.
- Tessler, M. H., & Goodman, N. D. (2019). The language of generalization. *Psychological Review*, 126(3), 395.
- Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, 38(4), 599–637.
- Wetherick, N. E., & Gilhooly, K. J. (1990). Syllogistic reasoning: Effects of premise order. *Lines of Thinking*, 1, 99–108.

Appendix: Pragmatic interpretation of premises

Understanding language often involves reasoning not only about the literal meaning of what was said but about why an interlocutor (the speaker) would bother to say what they said, which can lead to enriched interpretations beyond the literal meanings. We can formalize this reasoning through a series of Bayesian models that recursively reason about one another, as developed in the RSA framework (Frank & Goodman, 2012; Goodman & Frank, 2016; Scontras et al., 2018). Concretely, we can describe a pragmatic listener L_1 who reasons about a speaker S_1 who produces utterances to convey information to the literal listener (Eq. 2):

$$L_1(s | u_1, u_2) \propto P(s) \cdot S_1(u_1, u_2 | s) \quad (\text{A.1})$$

$$S_1(u_1, u_2 | s) \propto \exp[\alpha_1 \cdot \ln L_0(s | u_1, u_2)] \quad (\text{A.2})$$

The definition of the pragmatic listener (Eq. A.1) mirrors that of the literal listener (Eq. 1) except in that the likelihood is defined as the probability that a speaker S_1 would produce utterances (premises) u_1 and u_2 given a state s . Following standard practice in RSA modeling, the speaker is a soft-max rational agent (with degree of rationality α_1) who produces utterances in order to convey information about the state s to the literal listener (formalized by the negative surprisal of state given the utterances: $\ln L_0(s | u_1, u_2)$).

Alternative utterances. The proportionality in Eq. A.2 implies a normalization over a set of alternative utterances (i.e., the set of sentences that the speaker could have produced); in our case, the set of alternative utterances amounts to a set of alternative syllogistic premises that a speaker could have produced. There are at least three sets of alternative premises that are *a priori* plausible and we tested models using each of: (a) alternative premises that have the same ordering of terms (i.e., syllogisms of the same form, or “figure”), but whose quantified relation between the terms could be different (e.g., if the speaker says *All As are Bs/All Bs*

are *Cs*, the alternative set is the set of syllogisms of the form $Q_1 \text{As are Bs}/Q_2 \text{Bs are Cs}$ where $Q_1, Q_2 \in \{\text{all}, \text{some}, \text{not all}, \text{none}\}$; (b) alternative premises that differ either in the quantified relation between terms or the ordering of the terms, but not both (i.e., utterances with an edit distance of 1); (c) the maximal set of alternative premises in which the quantifier and the ordering of terms could be different.

Question under discussion. When deciding what to say, speakers may choose their utterances to convey information about a particular topic of relevance and listeners may assume that speakers had a particular topic of relevance in mind when interpreting the utterances heard. We formalize this topic as a QUD, which acts as a function that projects the state space onto a lower dimensional representation of the state space (Kao, Wu, Bergen, & Goodman, 2014; Roberts, 2004). Though QUDs may not be the explicit question asked of a person (e.g., Hawkins, Stuhlmüller, Degen, & Goodman, 2015), in the context of a syllogistic reasoning problem, the most natural QUD is the explicit question asked of participants: What relationship between *As* and *Cs* follows from the premises? (Tessler & Goodman, 2014). In particular, we considered the QUD when interpreting the premises of a syllogism (*As are Bs*, *Bs are Cs*) to be about the *A-C* relationship. In our model, this QUD function acts by transforming a three circle Venn diagram (*A*, *B*, *C*) into a two circle diagram (*A*, *C*): $Q(\{A, B, C\}) = \{A, C\}$.

The pragmatic listener who interprets the premises relative to this QUD is given by a set of equations:

$$L_1(s | u_1, u_2, Q) \propto P(s) \cdot S_1(u_1, u_2 | s, Q) \quad (\text{A.3})$$

$$S_1(u_1, u_2 | s, Q) \propto \exp[\alpha_1 \cdot \ln L_0(Q(s) | u_1, u_2, Q)] \quad (\text{A.4})$$

$$L_0(Q(s) | u_1, u_2) = \sum_{s' \in \mathcal{S}} \delta_{Q(s)=Q(s')} \cdot L_0(s' | u_1, u_2) \quad (\text{A.5})$$

In our explorations of the QUD model with various sets of alternative utterances, we did not find any substantial improvement in quantitative fit to the Ragni et al. (2019) data set.