Some arguments are probably valid: Syllogistic reasoning as communication

Michael H. Tessler, Noah D. Goodman
{mtessler, ngoodman}@stanford.edu
Department of Psychology, Stanford University

Abstract

We develop a computational-level theory of syllogistic reasoning which places reasoning at the intersection of communication and logic. The model considers reasoning over concrete situations and situations to be constructed probabilistically by sampling. We compare our model predictions with behavioral data from a recent meta-analysis. We show the flexibility of the model to account for reasoning behavior in a study of syllogisms using the generalized quantifiers *most* and *few*. We conclude by relating our model to two extant theories of syllogistic reasoning – Mental Models and Probability Heuristics – and discussing future directions.

Keywords: Reasoning, language understanding, probabilistic model

Consider for a moment that your friend tells you: "Everyone in my office has the flu and, you know, some people with the flu are out for weeks." Do you respond with "Everyone in your office has the flu." Do you respond with "Pardon me, there is no inference I can draw from what you just said." Or do you respond "I hope your officemates are not out for weeks and I hope you don't get sick either." The first response is true, but does not go beyond the premises; the second response attempts to go beyond the premises by strict classical logic, but fails; the final response goes beyond the premises, to offer a conclusion which is probably useful and probably true. This cartoon illustrates a critical dimension along which cognitive theories of reasoning differ: whether the core and ideal of reasoning is deductive validity or probabilistic support. A separate dimension concerns the extent to which principles of natural language—pragmatics and semantics—are necessary for understanding reasoning. In this paper we explore a theory of syllogistic reasoning inspired by recent advances in probabilistic semantics and pragmatics.

The form of the argument above resembles a syllogism: an argument with two quantifier expressions (premises) used to relate two properties (or terms) via a middle term. Fit into a formal syllogistic form, this argument would read:

All officemates are out with the flu
Some people out with the flu are out for weeks
Therefore, some officemates are out for weeks

The full space of syllogistic arguments is derived by shuffling the term-ordering ("All A are B" vs. "All B are A") and changing the quantifier (*all, some, none, not all*). Most syllogisms have no valid conclusion, i.e. there is no deductive relation between A & C determined by the premises. This is the case with the argument above. A recent meta-analysis of syllogistic reasoning tasks showed that over the population, accuracy on producing valid conclusions ranges from 90 % to 1% and ability to produce appropriate *no valid conclusion* responses ranges from 76% to 12% (Khemlani & Johnson-Laird, 2012): people are not good at drawing deductively valid conclusions.

Perhaps because of this divergence between human behavior and deductive logic, syllogistic reasoning has been a topic of considerable interest in cognitive psychology for over one hundred years (Störring, 1908), and before that in philosophy, dating back to Aristotle. In cognitive psychology we are interested in how people reason, and syllogisms lie at the intriguing intersection of natural and formal reasoning, of language and logic. They are undoubtedly logical; indeed, syllogisms were the first and only formal system of logic for millennia. At the same time, they use natural language quantifiers and invite natural language conclusions; precisely pinning down the meaning and use of quantifiers has been an ongoing area of inquiry since Aristotle (e.g. Horn, 1989).

Many theories of syllogistic reasoning take deduction as a given and try to explain errors as a matter of noise during cognition. These errors may arise from improper use of deductive rules or biased construction of logical models. At the same time, many other kinds of reasoning have been explained as probabilistic inference under uncertainty. Probability provides a natural description of a world in which you don't know how many people are in the hallway outside your door or whether or not the lion is going to start charging. We suggest that combining probabilistic reasoning with formal semantics and pragmatics of natural language leads to a useful combination of approaches, in which our knowledge describes distributions on possible situations and sentences naturally update these distributions with new information. In this formalism, deduction emerges as those arguments which are always true and syllogistic reasoning becomes a process of determining what is most probable, relevant, and informative.

The Probabilistic Questioner-Reasoner model

Our model begins with the intuition that people reason probabilistically about situations populated by objects with properties. To represent this type of richly structured situation, we must go beyond propositional logic and its probabilistic counterpart, Bayesian networks. We instead build our model using the probabilistic programing language Church (Goodman, Mansinghka, Roy, Bonawitz, & Tenenbaum, 2008), a kind of higher-order probabilistic logic in which it is natural to describe distributions over objects and their properties. For background and details on this form of model representation, see `http://probmods.org`.

Situations are composed of *n* objects:

```
(define objects (list 'o1 'o2 ... 'on))
```

(Ellipses indicate omissions for brevity, otherwise models are specified via runnable Church code.) Properties `A`, `B`, and `C` of these objects are represented by functions from objects to the property value. We assume properties are Boolean, and so property values can be `true` or `false`. We assume no *a priori* information about the meaning of the properties and thus they are determined independently:

```
(define A (mem (lambda (x) (flip br))))
(define B (mem (lambda (x) (flip br))))
(define C (mem (lambda (x) (flip br))))
```

Note that the operator `mem` memoizes these functions, so that a given object has the same value each time it is examined within a given situation, even though it is initially determined probabilistically (via `flip`). Previous probabilistic models (Oaksford & Chater, 1994) have invoked a principle of rarity from the observation that properties are relatively rare of objects in the world[1]. For us this simply means the base rate, `br`, of properties is small.

We interpret quantifiers as truth-functional operators, consistent with standard practice in formal semantics. A quantifier (e.g. *all*) is a function of two properties (e.g. *As* and *Bs*) which maps to a truth value by consulting the properties of the objects in the current situation. For instance:

```
(define all
  (lambda (A B)
    (all-true (map (lambda (x) (if (A x) (B x) true))
                   objects))))
```

Here the helper function `all-true` simply checks that all elements of a list are true, i.e. that all the *As* are indeed *Bs*. The function `map` applies the given function —`(lambda ...)`— to each element of the list `objects`. Similarly we can define `some`, `no`, `not-all` to have their standard meanings. For a first test of the model, we assume sets are non-empty, i.e. *all* and *none* cannot be trivially true.

The key observation to connect these truth-functional meanings of quantifier expressions to probability distributions over situations is that an expression which assigns a Boolean value to each situation can be used for probabilistic conditioning. That is, these quantifier expressions can be used to update a prior belief distribution over worlds into a posterior belief distribution. For syllogistic reasoning we are interested not in the posterior distribution over situations directly, but the distribution on true conclusions that these situations imply. In Church this looks like:

```
(query
 (define objects (list 'o1 'o2 ... 'on))
 ...define A,B,C...
 ...define all, some, no, notall...
 (define conclusion (conclusion-prior))

 conclusion

 (and (conclusion A C)
      (premise-one A B)
      (premise-two B C)))
```

[1]This article is an article and it's about reasoning, but it's not a cat, and it's not a car, nor an elephant nor the color red. In fact, there's a very large number of things which this article is not.

The first arguments to a query function are a generative model: definitions or the background knowledge with which a reasoning agent is endowed. Definitions with which a prior is stipulated (e.g. `conclusion`) denote concepts over which the agent has uncertainty . The second argument, called the *query expression*, is the aspect of the computation about which we are interested; it is what we want to know. The final argument, called the *conditioner*, is the information with which we update our beliefs; it is what we know.

We assume that the prior distribution over conclusions (and premises, below) is uniform.

Recursion and pragmatics

We have suggested viewing syllogistic reasoning as a case of communication, and this in turn suggests that reasoning should go beyond the semantics of language, to its pragmatics. Natural language pragmatics could enter in two places in the above model: premise interpretation and conclusion production. We address pragmatic production, or choice of conclusion, first.

Following the Rational Speech Act theory (Goodman & Stuhlmüller, 2013; Frank & Goodman, 2012) we imagine a reasoner who chooses the conclusions which are more likely to convey information about which only the reasoner has access—in this case, the premises. That is, we imagine that the reasoner treats the conclusion as a message chosen for a listener who lacks the premises (or perhaps lacks only the ability to reason from premises to conclusions). This listener can be thought of concretely as the (rather naive) person evaluating the participant's responses. In the original Rational Speech Act model, the listener hears an utterance and tries to reconstruct the situation observed by the speaker. In our model, the *conclusion-listener* hears a conclusion and tries to reconstruct the premises with which the reasoner was presented. This distinction between situation and premise reconstruction turns out to be important as the reasoner does not observe a situation directly but rather only constructs situations consistent with the premises: our reasoner attempts to convey her *knowledge state*, not her guess about the situation.

Pragmatics may enter into syllogistic reasoning also in the interpretation of the premises by the reasoner. It is known that standalone Gricean implicatures (for instance "some" implies "not all") do a poor job of accounting for reasoning with syllogisms (M. J. Roberts, Newstead, & Griggs, 2001). Our preliminary analyses with a standard Gricean-listener model in this framework were consistent with this account. However, a reasoner may consider the premises in a wider, conversational setting: she may ask herself why the experimenter chose to give these premises, as opposed to alternative utterances. For this we consider the question she believes to be at issue in the "conversation"—the Question Under Discussion, or QUD (C. Roberts, 2004). In a syllogistic conversation, we take the QUD to be "what is the relationship between A & C (the end terms)?", often the question with which the subject is explicitly presented. In this context, pragmatic inferences may differ from the standard ones; for instance, "Some A are B" may not lead to a "Not all A are B" implicature if "All A are B" wouldn't provide additional information about the A-C conclusion. The A-C QUD is naturally captured by a `questioner` who considers only the conclusion the `reasoner` would draw about A & C (not her inferences about the whole situation, which includes B).

We can combine the above intuitions about pragmatic production and comprehension into a model in which `reasoner` and `questioner` jointly reason about each other. Critically each agent reasons about the other at recursive depth `Q_depth` of comprehension and `R_depth` of production:

```
(define (questioner conclusion Q_depth R_depth)
  (query
   (define premise-one (premise-prior))
   (define premise-two (premise-prior))

   (list premise-one premise-two)

   (equal? conclusion (softmax (reasoner premise-one
        premise-two Q_depth R_depth) alpha))))

(define (reasoner premise-one premise-two Q_depth R_depth)
  (query
   (define objects (list 'o1 'o2 ... 'on))
   ...define A,B,C...
   ...define all, some, no, notall...
   (define conclusion (conclusion-prior))

   conclusion

   (and (if (= Q_depth 0)
            (and (premise-one A B)
                 (premise-two B C))
            (equal? (list premise-one premise-two)
                    (softmax (questioner conclusion
                    (- Q_depth 1) R_depth) alpha)))
        (if (= R_depth 0)
            (conclusion A C)
            (equal? (list premise-one premise-two)
                    (softmax (questioner conclusion
                         Q_depth (- 1 R_depth))
                         alpha))))))
```

The `reasoner` and `questioner` functions produce a distribution over conclusions and premises[2], respectively. Since we take these functions to represent actual persons in a communicative setting, we take conclusions or premises to be selected from these distributions according to a Luce choice, or `softmax`, decision rule with a parameter `alpha` that denotes the degree to which utterances are chosen optimally (Luce, 1959). This takes the distribution, raises it to a power `alpha` and renormalizes. As `R_depth` increases, the conclusion becomes more informative with respect to the possible premises it could have been drawn from. As `Q_depth` increases, the premises becomes more informative with respect to the possible conclusions. In this article, we explore these two parameters independently, referring to the model as the Probabilistic Questioner-Reasoner model or $\text{PQR}_{q,r}$, where q and r refer to respective depths. When q and r are 0, the model collapses to produce the Pr(conclusion | premises) [3].

Using this framework, we can investigate the complementary influences of pragmatic interpretation of premises and production of conclusions.

[2]As a first pass, we consider the alternative premises generated by `premise-prior` to be the set of all premises of the same term orderings, i.e. all premises of the same *figure*. That is, we consider alternative quantifiers, keeping the structure of the sentence fixed.

[3]This 0,0 model will be referred to the literal reasoner as well

Results

To test our model we used data from the meta-analysis of syllogistic reasoning tasks presented by Chater and Oaksford (1999). These data were compiled from five studies on syllogistic reasoning, completed from 1978-1984. The data include percentage response for conclusions that contain each of the 4 quantifiers as well as for "No Valid Conclusion" (for simplicity, we don't consider "No Valid Conclusion" in the current model). Some studies in the meta-analysis asked participants to draw conclusions which were restricted to the classical ordering of terms (C-A) while others allowed conclusions in either direction (A-C or C-A). To accommodate this, we allowed our model to draw conclusions in either order and collapsed responses across these two orderings to compare it to the data set. The three parameters of the model, `n_objects`, `br`, and `alpha`, were set to 6, 0.25, and 2, respectively, to give sensible[4] results.

Qualitative results

For each model, we report the total number of syllogisms for which the model's modal response is the same as in the meta-analysis. This is a qualitative assessment of the fit. The table below shows the number of modal responses for which the model matched the data (columns "matches"). We separate these into valid and invalid syllogisms[5]. The total numbers of valid and invalid syllogisms are 24 and 40, respectively.

Model	matches_{valid}	$\text{matches}_{invalid}$	r_{valid}	$r_{invalid}$
Prior	5	24	-.46	.42
$\text{PQR}_{0,0}$	17	20	-.21	.62
$\text{PQR}_{1,0}$	17	24	.82	.63
$\text{PQR}_{0,1}$	21	21	.69	.72
$\text{PQR}_{1,1}$	20	21	.85	.58

As a baseline, we first examined the distribution of conclusions conditioned only on the truth of the conclusion (what we refer to as the "Prior") to see if it alone accounted for human reasoning patterns. It did not (Figure 1, column 1). Since *not all* is the most likely conclusion to be true, the Prior matches only the syllogisms with a *not all* modal response. The literal reasoner model, $\text{PQR}_{0,0}$, matches the modal response on 37 of the 64 syllogisms. The 29 syllogisms for which *not all* was the modal response are qualitatively unaffected. The model also matches 8 syllogisms for which *some* and *none* are favored (Figure 1, column 2, see e.g. [2]). Thus, taking the premises into account in the most straightforward way improves performance, but not greatly so.

Introducing pragmatics into the model, we consider first production—$\text{PQR}_{0,1}$. The model now distinguishes between equally valid conclusions on the basis of informativity (e.g. Figure 1 [1], column 4). The model selects not only conclu-

[4]`n_objects` = 6 keeps the number of distinct, possible objects in a situation to a minimum while `br` = 0.25 is in accord with the rarity assumption.

[5]Since the response format in the meta-analysis varied across studies, the number of valid syllogisms was also not the same. Here we count as valid only the syllogisms that would have been considered valid in all studies.

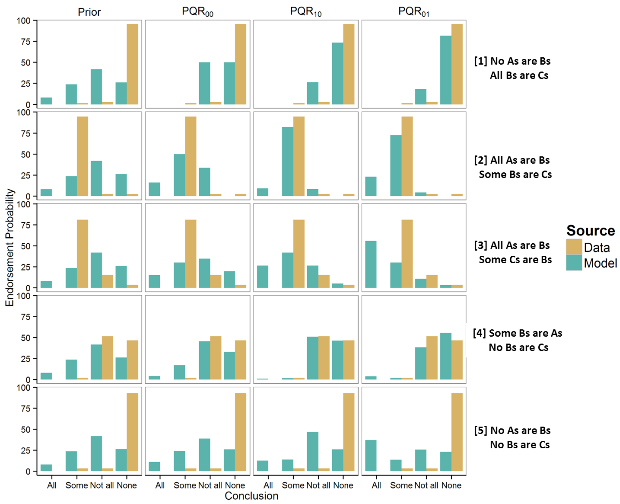

Figure 1: Five example syllogisms. [1] Literal reasoner has no preference among equally valid conclusions; the symmetry is broken by the pragmatics models which include informativity either in interpretation or in production. [2] Literal reasoner alone captures the modal response and pragmatics enriches the quantitative fit. [3] Relatively uninformative premises suggest *Some* is the most likely interpretation. [4] Models are able to capture multiple preferred conclusions. [5] Models do poorly in matching subjects' responses in an underconstrained, invalid syllogism.

sions likely to be true, but also informative, and now matches 42 out of 64 modal responses. In addition to capturing many of the modal responses, the model is able to accommodate more than one plausible conclusion. [4] in Figure 1 is one such example. This is a syllogism with a valid conclusion, but one which people find difficult to draw. The literal reasoner model tells us why: in many of the possible situations in which the premises are true, a *none* conclusion is true. In addition, *none* is more informative than the valid conclusion—*not all*—and so the Questioner-Reasoner Models strengthen the plausible but invalid *none*.

Conversational pragmatics can also enrich the meaning of the premises given to the reasoner— $\text{PQR}_{1,0}$— by considering "why has the questioner produced these premises given that she may have produced a variety of others?" $\text{PQR}_{1,0}$ maximally-prefers the modal response of subjects for 41 out of 64 syllogisms. As well, it picks up on some of the very complex phenomena present in syllogistic reasoning. Example [3] is one such case. The premises considered literally are relatively uninformative. The literal reasoner is very similar to the Prior (columns 1 & 2). In this setup, $\text{PQR}_{0,1}$ selects the most informative answer – *all*. However, $\text{PQR}_{1,0}$ considers the question "why would the questioner have given these premises" and concludes *some*, consistent with human responses.

Though this is all encouraging qualitative data, there a number of syllogisms for which reasoning patterns are not accounted by PQR. Many of these are syllogisms use two negative quantifiers (*not all* or *none*) as the premises. In these problems the predictions of the literal reasoner do not differ appreciably from the predictions of the Prior (Figure 1 [5]), because the rarity prior assumes most relations will be false to begin with. This perhaps indicates a mismatch between the prior we have used and people's prior beliefs when representing arbitrary sets.

Model fit

To assess our models' quantitative fits we examine correlations across all 256 data points (64 syllogisms x 4 conclusions), shown in Figure 2. The Prior's predictions are the same for all syllogisms and the overall fit is poor (r = 0.36). After conditioning on the truth of the premises, the model is able to make graded responses. These responses are a reflection of the types of situations consistent with the premises. The overall correlation is appreciably higher (r = 0.64). Among valid conclusions, however, (squares in Figure 2) the fit is terrible (r = -0.20 for valid conclusions only). This is a direct consequence of the reasoner's literalness: the model has no preference among multiple valid conclusions, since each is true in every situation.

This symmetry is broken by using pragmatic reasoning (Figure 2, columns 3-5). The reasoner who interprets the premises as coming from a pragmatic questioner—$\text{PQR}_{1,0}$—provides so far the best fit to the data (r = 0.75). The model is now able to make graded responses among valid conclusions (r = 0.62 for valid conclusions only). The reasoner who takes the premises at face value but who produces conclusions to be informative—$\text{PQR}_{0,1}$—has also a better fit than the literal model (r = 0.69) and the fit among valid conclusions is even stronger (r = 0.82 for valid conclusions only).

$\text{PQR}_{1,1}$—which reasons pragmatically on both the production and the interpretation—does not model the data overall any better than the individual processes. This joint model selects the modal response on 41 out of the 64 syllogisms and provides a worse fit than either of the individual PQR models (r = 0.66). However, among valid conclusions the correlation is highest (r = 0.85). This mismatch is puzzling and likely due to the particular way in which the information between the two inference processes is shared. We leave for later work the proper way of combining information from the two loci of pragmatic inference.

Most and few

Our model is based on a truth-functional semantics and as such, it is able to accommodate any quantified sentence with a truth-functional meaning. The meaning of generalized quantifiers like "most" and "few" is a topic of debate in formal semantics, but they are often simply modeled as a thresholded function. As a first test of the generality of the model, we define most and few by a threshold of 0.5 such that "most As are Bs" is true if more than half of the As are Bs. Once we have added these lexical items, our PQR models extend naturally. We compare our model predictions to two studies car-

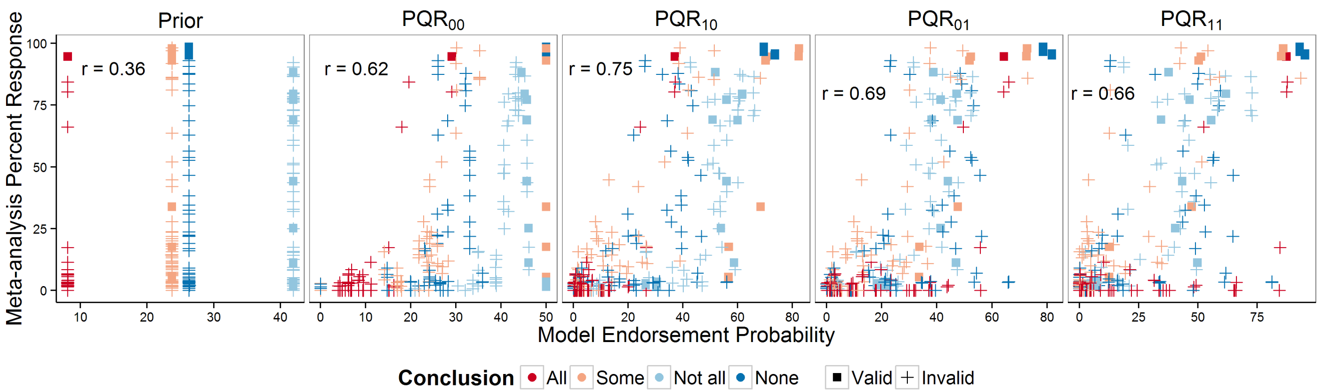

Figure 2: Human subject percentage endorsement vs. model fits. Columns (from L to R): predictions based only on the prior, "literal" probabilistic reasoner, and Questioner-Reasoner models. Subscripts denote depth of recursion for interpretation of premises and production of conclusions, respectively.

ried out by Chater and Oaksford (1999) on syllogisms using the generalized quantifiers *most* and *few* e.g. *Most artists are beekeepers. Few chemists are beekeepers.*. Participants were told to indicate which, if any, of the four quantifier conclusions followed from the premises and were allowed to select multiple options. The set of syllogisms was divided into two experiments to avoid subject fatigue.

We find good correspondence between the experimental data and the model, even without further parameter fitting[6] (Figure 3). In Experiment 1, the quantifiers *all*, *most*, *few*, and *not all* were used. In Experiment 2, the quantifiers *most*, *few*, *some*, and *none* were used. Note again the total number of syllogisms in an experiment is 64.

Model	matches_{Exp_1}	matches_{Exp_2}	r_{Exp_1}	r_{Exp_2}
$PQR_{0,0}$	45	34	.79	.63
$PQR_{1,0}$	48	43	.81	.63
$PQR_{0,1}$	50	44	.76	.60

The fit is appreciably better for Experiment 1 than for Experiment 2, and the same was true for the Probability Heuristics Model (r = 0.94 vs r = 0.63). Overall, the proportion of *no valid conclusion* responses in the experimental data, which we do not model, was much higher in Experiment 2 than in Experiment 1. Hence, this data set may well contain more noise in the measures we model than the others discussed in this article.

Discussion

The inspiration for our PQR model comes from the idea that syllogistic reasoning cannot be disentangled from language understanding. Natural language semantics alone seems to be insufficient to explain the variability in reasoning, however. We shown that a combination of semantics and conversational pragmatics provides insight into how people reason with syllogistic arguments.

[6]We maintain the `n_objects` parameter of 6, though the words "most" and "few" might pragmatically infer sets of larger size.

A recent meta-analysis carved the space of reasoning theories into three partitions: those based on models or diagrammatic reasoning, those based on formal logical rules, and those based on heuristics (Khemlani & Johnson-Laird, 2012). We see the space slightly differently. In one dimension, theories may be based on directly applying derivation rules—be they heuristic or logical—or they may be based on constructing concrete representations or models. In another dimension, theories may be considered fundamentally probabilistic or deterministic. This theoretical partitioning places the Probabilistic Questioner-Reasoner model in a previously unexplored quadrant of the two-dimensional theoretical space described: we consider probabilistic reasoning over concrete situations.

The Mental Models Theory (MMT) was offered to capture the intuition that people are able to reason about sets of things explicitly and with respect to context by constructing mental representations of individuals over which to reason. The situations described in PQR are analogous to mental models. MMT, however, leaves to complex heuristics the problem of which models come into existence. In PQR the situations follow from the semantic assumptions and fundamental laws of probability with no further assumptions; pragmatics then modifies this well-defined reasoning process.

Chater and Oaksford (1999) introduced the Probability Heuristic Model (PHM) which first derives a set of probabilistic rules for syllogistic reasoning; to account for informativity and other effects the PHM then augments these probabilistic rules with a complex set of heuristics (for example, informative-conclusion heuristics). Our PQR differs in two respects. First, the probabilistic "rules" emerge naturally from the semantics of quantifiers by reasoning about situations. Second, we further strengthen inferences by employing previously-proposed formalisms for probabilistic reasoning. This gives rise to many of the same effects, such as informativity, without postulating heuristics *de novo*.

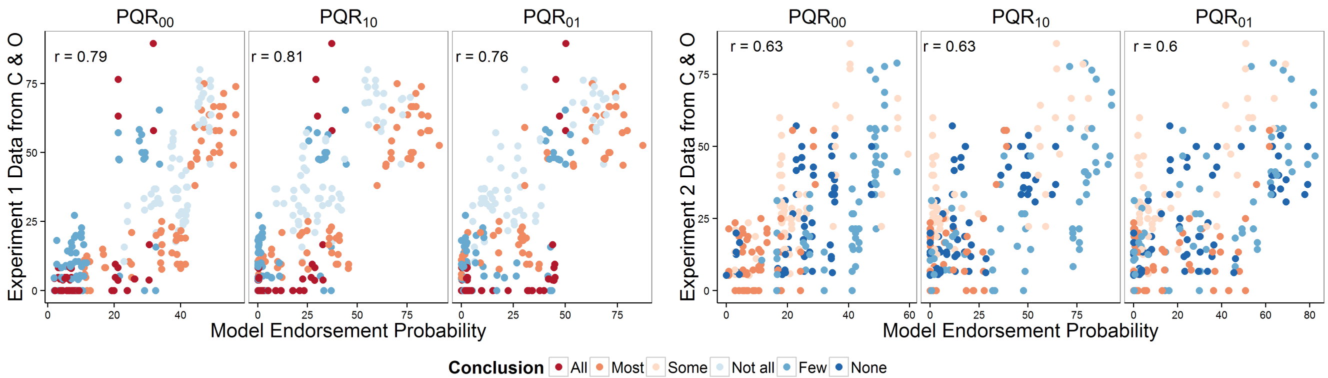

Figure 3: Human subject percentage endorsement vs. model fits for 2 experiments using generalized quantifiers. Experiment 1 (left) used the quantifiers {*all*, *most*, *few*, *not all*}. Experiment 2 (right) used the quantifiers {*most*, *few*, *some*, *not all*}.

This is early work and we have found promising evidence, both qualitative and quantitative, that this framework will allow for a more explicit understanding of syllogistic reasoning. $\text{PQR}_{1,0}$ and $\text{PQR}_{0,1}$ account for much of the data and in non-overlapping ways. The number of syllogisms whose modal response was not predicted by either model is a mere 9 (i.e. 55 out of 64 syllogisms were accounted for by one of the models). However, $\text{PQR}_{1,1}$ in its current formulation is not simply a sum of its parts. Evidence from the individual models suggests some complex interaction between pragmatic interpretation of premises and production of conclusions is at work in human reasoning behavior.

A major virtue of the PQR framework is that it extends naturally to incorporate any terms for which a truth-functional semantics can be given. For instance, we tested the model on *most* and *few* using the simplest, most standard semantics (most is more than half, etc). It is likely that these quantifiers actually have more complex semantics, but even so we accounted for a significant fraction of the data.

In this framework, a syllogism is read as an argument given as a part of discourse between people. Indeed, this is how syllogisms were used in the time of Aristotle and in the long tradition of scholastic philosophers since. Fundamentally, syllogisms are a tool used to convince others. The results of the Questioner-Reasoner models recast the ancient idea that human reasoning behavior in the syllogistic task is as much reason as it is human. Gauging degrees of truth or plausibility alone is not sufficient. An agent needs to be posited at the other end of the line so that a conclusion makes sense; so that an argument may convince!

References

Chater, N., & Oaksford, M. (1999). The Probability Heuristics Model of Syllogistic Reasoning. *Cognitive psychology*, *258*, 191–258.

Frank, M. C., & Goodman, N. D. (2012). Quantifying pragmatic inference in language games. *Science*, *336*, 1–9.

Goodman, N. D., Mansinghka, V. K., Roy, D. M., Bonawitz, K., & Tenenbaum, J. B. (2008). Church : a language for generative models. *Uncertainty in Artificial Intelligence*.

Goodman, N. D., & Stuhlmüller, A. (2013, January). Knowledge and implicature: modeling language understanding as social cognition. *Topics in cognitive science*, *5*(1), 173–84.

Horn, L. R. (1989). *A natural history of negation*. Chicago: University of Chicago.

Khemlani, S., & Johnson-Laird, P. N. (2012, May). Theories of the syllogism: A meta-analysis. *Psychological bulletin*, *138*(3), 427–57.

Luce, R. D. (1959). *Individual choice behavior*. New York, NY: Wiley.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*(4), 608–631.

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

Roberts, C. (2004). Information structure in discourse. *Semanatics and Pramatics*(5), 1-69.

Roberts, M. J., Newstead, S. E., & Griggs, R. a. (2001, May). Quantifier interpretation and syllogistic reasoning. *Thinking & Reasoning*, *7*(2), 173–204.

Störring, G. (1908). Experimentelle untersuchungen uber einfache schlussprozesse. *Arch. f. d. ges. Psychol*, 1-127.