Reasoning About the Sizes of Sets: Progress, Problems, and Prospects

Lawrence S. Moss[1], **Charlotte Raty**[2]
[1] Mathematics Department, Indiana University, Bloomington, IN 47405, USA
[2] Mathematics Department, Swarthmore College, 500 College Ave, Swarthmore, PA 19081, USA

lmoss@indiana.edu, craty1@swarthmore.edu

Abstract

We discuss what is known about logical reasoning concerning the sizes of sets, including expressions like *there are at least as many x as y*, *there are more x than y*, and *most x are y*. It turns out that reasoning with expressions like this can be done efficiently, that formal proofs can be obtained which do not employ translation to standard logic, and that counter-models can also be generated. The paper also contains a new result, a completeness theorem for syllogistic reasoning involving the sentences in our fragment, and adding sentential negation. So we are not done with the project of getting a complete logics for reasoning about sizes of sets. At the same time, there are some open questions. We show one implementation. We mention briefly some very new work which allows one to do logical reasoning on sentences as they come (not from a toy grammar), but at some cost. Finally, we discuss connections to cognitive science, many of which are waiting to be made.

1 Introduction

This workshop is concerned with Bridging the Gap between Human and Automated Reasoning. The specific contribution here might be called *automated reasoning pertaining to human reasoning, but done in a setting that deviates from the normative frameworks in logic*. We are interested in simple forms of reasoning about the sizes of sets. Such reasoning is very common in "real-world" settings, certainly more common than hard-core logical or mathematical reasoning of the type that theorem-provers handle so well.

For example, if someone says that *there are more ducks than geese in the pond* and then goes on to say that *there are more fish then ducks in the pond*, it follows that *there are more fish than geese in the pond*. This is the kind of reasoning we are after in this study. It is natural, yet it is not expressible in first-order logic. This means that *no existing first-order* theorem prover could carry out our simple reasoning, simply because it can't be represented in the logic.

Now one might counter that our simple can be handled in *second-order logic*, or in first-order logic with an extra theory of natural numbers. We feel that this is misguided for several reasons. First of all, first-order logic is already *undecidable*, and adding extra gadgetry cannot fix this. One of the primary lessons from cognitive science and AI to logic is that we should aim for *light* systems, since these are the ones that are the most cognitively plausible ones. And an additional point about numbers: the kind of "more than" reasoning that we are discussing has very little to do with counting, or at least its scope goes beyond counting. Consider mass nouns like *sand* and *water*. We can say that *there is more sand than water in the pond*. To do reasoning with assertions like this does not seem to require numbers. Finally, even if one had a logical system with first-order logic, and also with numbers, and implemented enough of the proof system to have a running implementation, would the resulting proofs be human-readable? We doubt that they would.

The natural logic program For a number of years, the first author has been working on a program of *natural logic*, building on the work of many others to be sure and also obtaining new results. Here are the main goals:

1. Show the aspects of natural language inference that can be modeled at all can be modeled using logical systems which are decidable.
2. To make connections to proof-theoretic semantics, and to psycholinguistic studies about human reasoning.
3. Whenever possible, to obtain *complete axiomatizations*, because the resulting logical systems are likely to be interesting, and also *complexity results*.
4. To implement the logics, and thus to have running systems that directly work in natural language, or at least in formal languages that are closer to natural language than to traditional logical calculi.
5. To re-think aspects of natural language semantics, putting inference at the center of the study rather than at the periphery.

The first, second, and third goals are rather close to those of the workshop, while the last two goals are farther. Most of the work on natural logic has been in the first and third goals, and so this paper is largely a report of ongoing work in those directions.

For more on natural logic, see [Abzianidze, 2015; van Benthem, 1986; Icard and Moss, 2014; McAllester and Givan, 1992; Moss, 2008; Pratt-Hartmann and Moss, 2009; Moss, 2015; 2016; Zamansky *et al.*, 2006].

It might come as a surprise, but reasoning about the sizes of sets does not by itself require a "big" logical system. We get an interesting logical system in the following way. Let's take *nouns* $p, q, x, y, z, \ldots$. And then we form *sentences* in the following very simple way: if x and y are nouns, then *all* x *are* y, *some* x *are* y, *there are at least as many* x *as* y, and *there are more* x *than* y are sentences. The first two forms of sentences are familiar from syllogistic logic, and indeed the logics in this paper are all *extended syllogistic logics*. We should emphasize that these logical systems do not have quantifiers; for that matter, they do not have propositional connectives $\wedge$, $\vee$, or $\neg$. The only sentences are the ones mentioned. In a sense, they are the cognitive module for reasoning about size comparison of sets, in the simplest possible setting.

Up until now, we have only given the syntax of a logical system. For the semantics, we consider *models* $\mathcal{M}$ consisting of a set M and subsets $[\![x]\!] \subseteq M$ for all nouns x. We then declare

$$\begin{array}{lll}
\mathcal{M} \models \textit{all } x \textit{ are } y & \text{iff} & [\![x]\!] \subseteq [\![y]\!] \\
\mathcal{M} \models \textit{some } x \textit{ are } y & \text{iff} & [\![x]\!] \cap [\![y]\!] \neq \emptyset \\
\mathcal{M} \models \textit{there are at least as many } x \textit{ as } y & \text{iff} & |[\![x]\!]| \geq |[\![y]\!]| \\
\mathcal{M} \models \textit{there are more } x \textit{ than } y & \text{iff} & |[\![x]\!]| > |[\![y]\!]|
\end{array}$$

Then we ask some very traditional questions. For a set Γ of sentences in this language, and for another sentence φ, we say that $\Gamma \models \varphi$ if every model $\mathcal{M}$ which satisfies all sentences in Γ also satisfies φ. This is the notion of *semantic consequence* from logic. We are especially interested in a refined notion, where we require that every *finite* model $\mathcal{M}$ which satisfies all sentences in Γ also satisfies φ. There is a difference, and we shall explore it below. For now, note that if $\mathcal{M}$ is finite, the we have the following inference: on the assumption that *all* x *are* y and also that *there are at least as many* x *as* y, it follows that (x and y are the same set, and hence) *all* y *are* x.

Again, the main reason to restrict to finite models is that this could be more relevant to natural reasoning.

Up until now, everything we have said concerns the semantics. In addition, there is a proof system. The relation between proof-theoretic and model-theoretic reasoning is extremely relevant to all studies of logic and cognitive science. This paper studies what might be taken as the classical approach, but we feel that this approach is worth re-thinking in future studies.

The proof system for this logic is a fairly standard system, in the natural deduction style. The rules are given in Figure 1, the part above the line. We have abbreviated all of the sentences in the logic, in the following way: $\forall(p, q)$ stands for *all* p *are* q, $\exists(p, q)$ stands for *some* p *are* q, and $\exists^{\geq}(p, q)$ stands for *there are at least as many* p *as* q, and $\exists^{>}(p, q)$ stands for *there are more* p *than* q.

For the most part, the rules in the system are evidently sound. This adds to the naturality of the system. The first few rules in are a complete syllogistic logic for sentences *all* x *are* y and *some* x *are* y [Moss, 2008]. The only place so far where the finiteness assumption is in the rule (CARD-MIX), and we have discussed this already.

The logic does not have *reductio ad absurdum*, but it has the related principle of *ex falso quodlibet*.

Adding set complements We obtain a more interesting system by adding more features to this logic. The one which we wish to explore here is *set complement*. For a noun x, we read $\overline{x}$ as "non-x". We allow nouns to be complemented. We identify the double complement $\overline{\overline{x}}$ with x itself in the syntax. And in the semantics, we take $[\![\overline{x}]\!]$ to be the complement set $M \setminus [\![x]\!]$.

The (HALF) rule, and other rules. If there are at least as many p as non-p, then the p's are at least half of the universe. So if there are at most as many q as non-q, then q's have at most half of the elements in the unvierse. Thus there are at least as many p as q. This is the content of the (HALF) rule.

The (MAJ) rule is the most complicated (and unexpected) rule in the system. It says that if there are at least as many p as non-p, and if there are at least as many q as non-q, and if some non-p is a non-q, then some p is a q. Why is this sound? The first two premises imply that they ps and q are each at most half of everything. Suppose towards a contradiction that no p are q. Then by size considerations, we see that the ps and qs are complementary halves. And so we contradict the last assumption. Now this form of reasoning, using *reductio* is not available in the logic. (Indeed, having *ex falso* rather than *reductio* makes the proof search which we describe below much more efficient.) So we need to take (MAJ) as a rule on its own rather than derive it from other principles. But then the rule is so complicated, it would be interesting to know whether educated "people on the street" could even understand it the same way that they naturally understand (BARBARA) or (DARII).

Theorem 1.1. [Moss, 2016] The logic in Figure 1 is sound and complete: $\Gamma \vdash \varphi$ iff $\Gamma \models \varphi$. Moreover, for finite $\Gamma \cup \{\varphi\}$, it is decidable in time polynomial in the length of $\Gamma \cup \{\varphi\}$ whether or not $\Gamma \vdash \varphi$.

1.1 Implementation

The logic has been implemented in CoCalc. (This language is close to Python and runs in Jupyter notebooks.) This implementation is on the cloud, and the first author shares it. See `https://cocalc.com/`. Other syllogistic logics have been implemented in Haskell.

For example, one may enter in the CoCalc implementation the following.

```
assumptions= ['All non-a are b',
'There are more c than non-b',
'There are more non-c than non-b',
'There are at least as many non-d as
d',
'There are at least as many c as
non-c',
```

$$\frac{}{\forall(p,p)}\ (\text{AXIOM}) \qquad \frac{\forall(n,p) \quad \forall(p,q)}{\forall(n,q)}\ (\text{BARBARA}) \qquad \frac{\exists(p,q)}{\exists(p,p)}\ (\text{SOME})$$

$$\frac{\exists(q,p)}{\exists(p,q)}\ (\text{CONVERSION}) \qquad \frac{\exists(p,n) \quad \forall(n,q)}{\exists(p,q)}\ (\text{DARII}) \qquad \frac{\forall(p,q) \quad \exists^{\geq}(p,q)}{\forall(q,p)}\ (\text{CARD-MIX})$$

$$\frac{\forall(p,q)}{\exists^{\geq}(q,p)}\ (\text{SUBSET-SIZE}) \qquad \frac{\exists^{\geq}(n,p) \quad \exists^{\geq}(p,q)}{\exists^{\geq}(n,q)}\ (\text{CARD-TRANS}) \qquad \frac{\exists(p,p) \quad \exists^{\geq}(q,p)}{\exists(q,q)}\ (\text{CARD-}\exists)$$

$$\frac{\exists^{>}(p,q)}{\exists^{\geq}(p,q)}\ (\text{MORE-AT LEAST}) \qquad \frac{\exists^{>}(n,p) \quad \exists^{\geq}(p,q)}{\exists^{>}(n,q)}\ (\text{MORE-LEFT}) \qquad \frac{\exists^{\geq}(n,p) \quad \exists^{>}(p,q)}{\exists^{>}(n,q)}\ (\text{MORE-RIGHT})$$

$$\frac{\exists^{\geq}(p,q) \quad \exists^{>}(q,p)}{\varphi}\ (\text{X})$$

$$\frac{\forall(p,\overline{p})}{\forall(p,q)}\ (\text{ZERO}) \qquad \frac{\forall(\overline{p},p)}{\forall(q,p)}\ (\text{ONE}) \qquad \frac{\forall(q,p) \quad \exists(p,\overline{q})}{\exists^{>}(p,q)}\ (\text{MORE})$$

$$\frac{\exists^{>}(p,q)}{\exists(p,\overline{q})}\ (\text{MORE-SOME}) \qquad \frac{\exists^{>}(q,p)}{\exists^{>}(\overline{p},\overline{q})}\ (\text{MORE-ANTI}) \qquad \frac{\forall(p,q)}{\forall(\overline{q},\overline{p})}\ (\text{ANTI})$$

$$\frac{\exists^{\geq}(p,q)}{\exists^{\geq}(\overline{q},\overline{p})}\ (\text{CARD-ANTI}) \qquad \frac{\exists(p,p) \quad \exists^{\geq}(q,\overline{q})}{\exists(q,q)}\ (\text{INT}) \qquad \frac{\exists^{\geq}(p,\overline{p}) \quad \exists^{\geq}(\overline{q},q)}{\exists^{\geq}(p,q)}\ (\text{HALF})$$

$$\frac{\exists^{>}(p,\overline{p}) \quad \exists^{\geq}(\overline{q},q)}{\exists^{>}(p,q)}\ (\text{STRICT HALF}) \qquad \frac{\exists^{\geq}(p,\overline{p}) \quad \exists^{\geq}(q,\overline{q}) \quad \exists(\overline{p},\overline{q})}{\exists(p,q)}\ (\text{MAJ})$$

Figure 1: Rules for the logic with *all p are q* (written $\forall(p,q)$ above), *some p are q* ($\exists(p,q)$), *there are at least as many p as q* ($\exists^{\geq}(p,q)$), and *there are more p than q* ($\exists^{>}(p,q)$). The rules below the line are for the system enlarged by taking set complements.

```
'There are at least as many non-d as
non-a']
conclusion = 'All a are non-c'
follows(assumptions,conclusion)
```

We are asking whether the conclusion follows from the assumptions. This particular set of assumptions is more complicated than what most people deal with in everyday life, even when they consider the sizes of sets. This is largely due to the set complements.

Please note that the implementation does not translate the assumptions into any other logical system. The interesting thing about logics in the syllogistic family is that if a sentence does not follow from a set of assumptions in the proof theory, then the proof search itself gives a counter-model. (This fact is not automatic, and it takes special arguments in each logic, but it generally does hold true.) The counter-model takes a little bit of extra work, but it is nothing like a search using a model-finder like Mace ([1]). Indeed, the counter-model search is a small add-on to the proof-search. In this example, the result appears instantly. We get back

[1] https://www.cs.unm.edu/ mccune/mace4/

```
The conclusion does not follow
Here is a counter-model.
We take the universe of the model
to be {0,1,2,3,4,5}
noun semantics
a     {2,3}
b     {0,1,4,5}
c     {0,2,3}
d     {}
```

So the system gives the semantics of a, b, c, and d as subsets of $\{0,\ldots,5\}$. Notice that the assumptions are true in the model which was found, but the conclusion is false.

Here is an example of a derivation found by our implementation. We ask whether the putative conclusion below really follows:

All non-x are x
Some non-y are z

There are more x than y

In this case, the conclusion does follow, and the system returns a proof.

$$\frac{\neg\exists^{\geq}(x,y)}{\exists^{>}(y,x)}\ (\text{NEGATE-AT LEAST}) \qquad \frac{\neg\exists^{>}(x,y)}{\exists^{\geq}(y,x)}\ (\text{NEGATE-MORE})$$

$$\frac{M(x,y)}{M(x,x)}\ (\text{MOST-LEFT}) \qquad \frac{M(x,y)}{M(y,y)}\ (\text{MOST-RIGHT})$$

$$\frac{M(x,y) \quad \exists^{\geq}(x,y)}{M(y,x)}\ (\text{MOST-AT LEAST}) \qquad \frac{\neg M(x,y) \quad M(y,x)}{\exists^{>}(x,y)}\ (\text{MOST-MORE})$$

$$\frac{}{\exists^{\geq}(x,x)}\ (\text{AT LEAST-AXIOM}) \qquad \frac{\neg M(x,y) \quad \exists^{\geq}(y,x)}{\neg M(y,x)}\ (\text{NOT-MOST-AT LEAST})$$

$$\frac{\neg M(x,x)}{\neg M(x,y)}\ (\text{NEGATE-MOST-LEFT}) \qquad \frac{\neg M(x,x)}{\neg M(y,x)}\ (\text{NEGATE-MOST-RIGHT})$$

$$\frac{\neg M(x,x)}{\exists^{\geq}(y,x)}\ (\text{NEGATE-MOST-CARD}) \qquad \frac{\neg M(x,x) \quad \exists^{\geq}(x,y)}{\neg M(y,y)}\ (\text{NEGATE-MOST-AT LEAST})$$

$$\frac{\neg M(x,x) \quad M(y,y)}{\exists^{>}(y,x)}\ (\text{NEGATE-MOST-MORE}) \qquad \frac{M(x,y) \quad \neg M(x,y)}{\varphi}\ (\text{X-MOST})$$

$$\frac{M(x_1,x_2) \quad \neg M(x_2,x_1) \quad M(x_2,x_3) \quad \neg M(x_3,x_2) \quad \cdots \quad M(x_{n-1},x_n) \quad \neg M(x_n,x_{n-1}) \quad M(x_n,x_1)}{M(x_1,x_n)}\ (\text{INF})$$

Figure 2: Most, at least, and more than, allowing sentential negation. Not shown: (CARD-TRANS) , (MORE-LEFT), (MORE RIGHT), (X), and (MORE-AT LEAST).

```
Here is a formal proof in our system:

1 All non-x are x      Assumption
2 All y are x      One 1
3 All non-x are x      Assumption
4 All non-y are x      One 3
5 Some non-y are z      Assumption
6 Some non-y are non-y      Some 5
7 Some non-y are x      Darii 4 6
8 Some x are non-y      Conversion 7
9 There are more x than y    More 2 8
```

Open question 1.2. If one drops the (MAJ) rule, is there a semantics which results in a sound and complete logic? If one adds (REDUCTIO) to the logic, then (MAJ) and several other rules become derivable. Would this new system be more cognitively plausible? Would it be possible to carry out the same kinds of questions that we ask concerning psycholingusitic studies (for example, those of [Geurts, 2005]) of syllogistic reasoning in order to craft a system which behaved like human reasoners?

Open question 1.3. The set complement operation here is understood as in classical logic: $[\![\overline{x}]\!] = M \setminus [\![x]\!]$. Occasionally people feel that this is too strong, that people would not say that every point belongs to either $[\![x]\!]$ or to $[\![\overline{x}]\!]$. One could try to do experiments to see how people actually behave regarding set complements in this setting, and then craft logical systems based on the data.

Variations The finiteness assumption is needed in the rules: (CARD-MIX), (MORE), (MORE-ANTI), (CARD-ANTI), and (STRICT-HALF); see [Moss and Topal, 2018]. If one wishes to interpret the same syntax on sets which are infinite, then the logic changes. We need to drop the rules which were mentioned just above, and in their place we need to add a few different rules. It is open to axiomatize the logic on sets which might be either finite or infinite.

It also is open to craft a logic that directly expresses size comparison about mass nouns. One place where we our logic is inappropriate is in the (INT) rule.

Additions Propositional connectives make complex sentences from sentences. We did not add the propositional connectives to this particular logic to keep the complexity low, and also to illustrate the idea that one should find special-purpose logics for topics like sizes of sets. However, one certainly could add in the connectives. The resulting logic has been explored, but we lack the space to expand on this.

$$\frac{\textit{most } x \textit{ are } y}{\textit{some } x \textit{ are } y}\ m_1 \qquad \frac{\textit{some } x \textit{ are } x}{\textit{most } x \textit{ are } x}\ m_2 \qquad \frac{\textit{most } x \textit{ are } y \quad \textit{all } y \textit{ are } z}{\textit{most } x \textit{ are } z}\ m_3$$

$$\frac{\textit{most } x \textit{ are } z \quad \textit{all } x \textit{ are } y \quad \textit{all } y \textit{ are } x}{\textit{most } y \textit{ are } z}\ m_4$$

$$\frac{\textit{all } y \textit{ are } x \quad \textit{all } x \textit{ are } z \quad \textit{most } z \textit{ are } y}{\textit{most } x \textit{ are } y}\ m_5 \qquad \frac{\textit{most } x \textit{ are } y \quad \textit{most } x \textit{ are } z}{\textit{some } y \textit{ are } z}\ m_6$$

Figure 3: Rules of the logical system for *All*, *Some*, and *Most*. We have not shown the rules (AXIOM), (BARBARA), (SOME), (CONVERSION), and (DARII). Also (m_6) is really just the first of an infinite sequence of rules.

2 Syllogistic Reasoning with the Word "Most"

We now turn to a different topic, reasoning with the word *most*, as in *Most dogs chase cats*. Settling on a semantics is already a difficult matter. We decided on *strict majority*:

$$\mathcal{M} \models \textit{most } x \textit{ are } y \quad \text{iff} \quad |[\![x]\!] \cap [\![y]\!]| > \tfrac{1}{2}|[\![x]\!]|$$

We abbreviate *most x are y* by $M(x, y)$. We add *sentential negation*. That is, the logic is rather trivial without sentential negation; see [Moss, 2008].

Of course, our semantics is a revisable choice. For just one example, we might decide to strengthen the definition of *most x are y* to mean that at least $\frac{2}{3}$ of the x's are also y's.

The logical system is shown in Figure 2. Actually, that figure shows the system needed when we add $\exists^{\geq}$. The main significant rule is the infinite scheme (INF) at the bottom. Extended syllogistic logics frequently have one or more rules which are highly non-trivial. One is tempted to say that it would take advanced training in mathematics to appreciate (INF); it would be surprising to find many real-life uses. Here is the understanding of the infinite scheme in Figure 2. Suppose that most x_1s are x_2s, but not conversely. Then there must be more x_2s than x_1s. So if we have a sequence of sets, say x_1, x_2, …, x_n as in the hypothesis of the rule, then we must have more x_2s than x_1s, more x_3s than x_2s, …, more x_ns than x_{n-1}s. We cannot have more x_1s than x_ns, because we would have a cycle of numbers in the $>$ ordering. So if most x_ns are x_1s, we must also have that most x_1s are x_ns.

Theorem 2.1 ([Lai *et al.*, 2016])**.** The complete logic for the logic containing sentences $M(x, y)$ and all propositional connectives is given by the rules (MOST-LEFT) and (MOST-RIGHT) in Figure 2, and adding the infinite scheme at the bottom of the figure. Moreover, the satisfiability problem for this logic is NP-complete.

The second author has extended this result[2].

Theorem 2.2 ([Raty, 2017])**.** The complete logic for sentences $M(x, y)$, $\exists^{\geq}(x, y)$, $\exists^{>}(x, y)$, and sentence negations is given by the rules (MOST-LEFT) and (MOST-RIGHT) in Figure 2, and the infinite scheme (INF). Moreover, the satisfiability problem for this logic is NP-complete.

We would also like to mention another completeness result in the area.

[2]Should this paper be accepted, and should there be enough space, we would like the eventual conference volume to contain a sketch of the proof.

Theorem 2.3 ([Endrullis and Moss, 2015])**.** The complete syllogistic logic of *All*, *Some*, and *Most* (but without sentential negation) is given by the logic in Figure 3, together with an infinite (but regular) scheme not shown in the figure. The inference problem for this logic is in polynomial time.

Remark 2.4. Many of the algorithmic questions on the logics of this paper follow from results in a different line of work, namely that of set constraints involving cardinality; see, for example [Kuncak *et al.*, 2006]. Here are some comparisons with our work. Work in our line involves logics rather than complexity results. Our work usually gives logics with lower complexity, matching the lower expressivity of our systems. The two areas call on different kind of combinatorial results.

Open question 2.5. The next big technical question in this line of work would be to combine the systems mentioned in Theorems 2.2 and 2.3. This seems like a challenging combinatorial problem.

Open question 2.6. We mentioned above that there are other options in thinking about *most*. But all of the options which we studied are in a way too "orthodox." One should investigate connections of *most* to *default logic* or to *reasoning via prototypes*, since these provide attractive ways of thinking about related phenomena.

Open question 2.7. [Rett, 2018] deals with the words *many*, *much*, *few*, and *little*. Of course there are many other publications on similar items, both in English and other languages. Are there convenient ways to directly incorporate linguistic treatments into logical systems? In the other direction, linguistics semanticists usually do not concern themselves with inference but rather stick to truth conditions. Is there any way that paying attention to inference, either in the standard sense or perhaps in some "non-monotonic" sense could help the semanticist?

3 Reasoning about sizes, without a fully specified grammar, without completeness, but allowing a broader set of input sentences

Up until now in this paper, we have worked with the most standard assumptions concerning logic and language. Logical languages come with a perfectly defined syntax and semantics. One of the things which we would like to do in natural logic is to question this set of assumptions, or at least to be aware of it at all times. What we want to do in this final section is to show how to loosen this assumption in connection with sizes of sets. It is based on [Hu and Moss, 2018].

This section calls on some knowledge of linguistic semantics. We are interested in *polarity marking*, as shown below:

More dogs$^{\downarrow}$ *than cats*$^{\uparrow}$ *walk*$^{\downarrow}$
Most$^{\uparrow}$ *dogs*$^{=}$ *who*$^{=}$ *every*$^{=}$ *cat*$^{=}$ *chased*$^{=}$ *cried*$^{\uparrow}$
Every dog$^{\downarrow}$ *scares*$^{\uparrow}$ *at least two*$^{\downarrow}$ *cats*$^{\uparrow}$

The $^{\uparrow}$ notation means that whenever we use the given sentence truthfully, if we replace the marked word w with another word which is "$\geq w$" in an appropriate sense (see below), then the resulting sentence will still be true. So we have a *semantic inference*. The $^{\downarrow}$ notation means the same thing, except that when we substitute using a word $\leq w$, we again preserve truth. Finally, the $^{=}$ notation means that we have neither property in general; in a valid semantic inference statement, we can only replace the word with itself rather than with something larger or smaller. We call $\uparrow$ and $\downarrow$ *polarity indicators*.

For example, suppose that we had a collection of background facts like *cats* $\leq$ *animals*, *beagles* $\leq$ *dogs*, *scares* $\leq$ *startles*, and one $\leq$ two. This kind of background fact could be read off from WordNet, thought of as a proxy for word learning by a child from her mother. In any case, our $^{\uparrow}$ and $^{\downarrow}$ notations on *Every dog*$^{\downarrow}$ *scares*$^{\uparrow}$ *at least two*$^{\downarrow}$ *cats*$^{\uparrow}$ would allow us to conclude *Every beagle startles at least one animal*. In general, $\uparrow$ notations permit the replacement of a "larger" word and $\downarrow$ notations permit the replacement of a smaller one.

The goal of the paper [Hu and Moss, 2018] is to provide a computational system to determine the notations $\uparrow, \downarrow, =$ on input text "in the wild". That means that the goal would be to take text from a published source or from the internet, or wherever, and to then accurately and automatically determine the polarity indicators. Then using a stock of background facts, we get a very simple "inference engine," suitable for carrying out a reasonable fraction of the humanly interesting inferences. The system would handle *monotonicity inferences* [Geurts, 2005; van Benthem, 1986]. Such a system would not be complete at all, because many forms of inference are not monotonicity inferences. For example, earlier in this paper we mentioned rules like (CARD-MIX) and (MAJ). Neither of these is a monotonicity rule. One could imagine taking our monotonicity inference engine and enriching it with natural logic rules; this is under active development.

We must emphasize that [Hu and Moss, 2018] is also not a complete success. The work there depends on having a correctly parsed representation of whatever sentence is under consideration, either as a premise of an argument or as the conclusion. And here is the rub: it is rather difficult to obtain semantically useable parses. We have wide-coverage parsers; that is, programs capable of training on data (probabilistically) and then giving structure to input sentences. And the parses of interest to us are those in a grammatical framework called *combinatory categorial grammar* (CCG). CCG is a descendant of categorial grammar (CG), and so it is lexicalized; that is, the grammatical principles are encoded in complex lexical types rather than in top-down phrase structure rules. From our point of view, this is a double-edged sword. On the one hand, CG and CCG connect syntax and semantics because combining constituents in the syntax is matched by function application in the semantics, or to combinators of various sorts. On the other hand, there is no real hope of writing a complete set of rules for logical systems whose syntax is so complicated as to require a grammar of this form. In effect, we give up on completeness in order to study a syntax that is better than a toy. For that matter, working with a wide-coverage parser for a framework like CCG means that some of the parses will be erroneous in the first place. And so the entire notion of deduction will have to be reconsidered, allowing for mistakes along the way.

As we mentioned, CCG parses are rather close to semantic representations, closer than the parses of other grammatical systems. And yet, they are still not always close enough to the semantics to be usable. So the project of inference from text "in the wild" is just beginning.

Open question 3.1. Would a monotonicity inference engine together with the simplest rules in this paper, together with logical principles like *reductio ad absurdum* be *in practice complete* for reasoning about the sizes of sets?

The question is vague, admittedly. In a sense, what it is asking is whether the various forms of work in this paper can be combined to give a full account of what people do when they reason about the sizes of sets.

4 Conclusion

Let us reiterate the main points in this paper.

First-order logic is both too big and too small to serve as the ideal logical language to represent human inference. It is too big because it is undecidable, and too small because it cannot represent many interesting phenomena that people reason about, such as size comparisons. So for this reason, we should look for other representational frameworks. Rather than dismiss the logical tradition entirely, we opt in this line of work to rehabilitate syllogistic logic, but enhanced with extra devices to talk about sizes of sets directly.

The logical systems that we propose are sound and complete, and frequently they are of polynomial complexity. When one adds propositional connectives, those connectives dominate the complexity. If one were to add boolean operations on the nouns and verbs, then again the extra expressive power due to the additions would dominate the complexity. But the basic logical systems of size comparison and of *most* are algorithmically manageable. And the systems have interesting principles, so those systems should be of independent interest.

The kind of work which we are describing is mainly a contribution to logic. But at every turn, we come face to face with interesting issues in semantics, in cognitive science, and in the study of human reasoning. We feel that people interested in the gap between automated and human reasoning might look at "small" phenomena, like reasoning about the sizes of sets, as a starting point for other studies. There is much to be done both on the side of automated reasoning, and on human reasoning.

References

[Abzianidze, 2015] Lasha Abzianidze. A tableau prover for natural logic and language. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 2492–2502, 2015.

[Endrullis and Moss, 2015] Jörg Endrullis and Lawrence S. Moss. Syllogistic logic with "most". In *Proceedings of WoLLIC*, pages 124–139. Springer LNCS, 2015.

[Geurts, 2005] Bart Geurts. Monotonicity and processing load. *Journal of Semantics*, 22(1):97–117, 2005.

[Hu and Moss, 2018] Hai Hu and Lawrence S. Moss. Polarity computations in flexible categorial grammar. In M. Nissim (et al), editor, *Proceedings of The Seventh Joint Conference on Lexical and Computational Semantics, *SEM 2018, New Orleans, Louisiana*, 2018.

[Icard and Moss, 2014] Thomas F. Icard and Lawrence S. Moss. Recent progress on monotonicity. *Linguistic Issues in Language Technology*, 9(7):167–194, 2014.

[Kuncak *et al.*, 2006] Viktor Kuncak, Huu Hai Nguyen, and Martin Rinard. Deciding Boolean Algebra with Presburger Arithmetic. *Journal of Automated Reasoning*, 36(3), 2006.

[Lai *et al.*, 2016] Tri Lai, Jörg Endrullis, and Lawrence S. Moss. Proportionality digraphs. *Proceedings of the American Mathematical Society*, 144:3701–3715, 2016.

[McAllester and Givan, 1992] David A. McAllester and Robert Givan. Natural language syntax and first-order inference. *Artificial Intelligence*, 56:1–20, 1992.

[Moss and Topal, 2018] Lawrence S. Moss and Selçuk Topal. Syllogistic logic with cardinality comparisons on infinite sets. *Review of Symbolic Logic*, 2018.

[Moss, 2008] Lawrence S. Moss. Completeness theorems for syllogistic fragments. In F. Hamm and S. Kepser, editors, *Logics for Linguistic Structures*, pages 143–173. Mouton de Gruyter, 2008.

[Moss, 2015] Lawrence S. Moss. Natural logic. In *Handbook of Contemporary Semantic Theory, Second Edition*, chapter 18. John Wiley & Sons, 2015.

[Moss, 2016] Lawrence S. Moss. Syllogistic logic with cardinality comparisons. In *J. Michael Dunn on Information Based Logics*, pages 391–415. Springer Outstanding Contributions to Logic, 2016.

[Pratt-Hartmann and Moss, 2009] Ian Pratt-Hartmann and Lawrence S. Moss. Logics for the relational syllogistic. *Review of Symbolic Logic*, 2(4):647–683, 2009.

[Raty, 2017] Charlotte Raty. Logic with "most" and cardinality comparisons. Unpublished ms., Indiana University, 2017.

[Rett, 2018] Jessica Rett. The semantics of many, much, few, and little. *Language and Linguistics Compass*, 12(1), 2018.

[van Benthem, 1986] Johan van Benthem. *Essays in Logical Semantics*, volume 29 of *Studies in Linguistics and Philosophy*. D. Reidel Publishing Co., Dordrecht, 1986.

[Zamansky *et al.*, 2006] Anna Zamansky, Nissim Francez, and Yoad Winter. A 'natural logic' inference system using the Lambek calculus. *J. Log. Lang. Inf.*, 15(3):273–295, 2006.