# The Good, the Bad, and the Difficult

Complexity in a Monotonicity-Grounded Natural Logic for Reasoning with Generalized Quantifiers

**MSc Thesis** *(Afstudeerscriptie)*

written by

**Jonathan Frederik Sippel**
(born June 12th, 1988 in Pfaffenhofen an der Ilm, Germany)

under the supervision of **Dr Jakub Szymanik**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

**MSc in Logic**

at the *Universiteit van Amsterdam.*

| **Date of the public defense:** | **Members of the Thesis Committee:** |
|---|---|
| *August 29, 2017* | Prof Dr Benedikt Löwe (Chair) |
| | Prof Dr Johan van Benthem |
| | Julian Schlöder |
| | Dr Tom Lentz |

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

# Abstract

We will present a natural logic (NQL) for reasoning with generalized quantifiers that aims to predict mean human success on syllogistic and related reasoning tasks. Natural logics provide inference rules that operate directly on natural language representations, thereby gaining flexibility and expressive power. NQL thereby proves to be more cognitively plausible than competing theories. We will further extend NQL to a natural language fragment that is concerned with quantifier iteration. The inference rules in NQL are assigned weights, corresponding to a measure of complexity of inferences – this weight assignment is motivated by semantic and psychological considerations. While the overarching goal is to align the complexity of sequences of inferences with the cognitive difficulty that reasoners encounter, we also aim to demonstrate that NQL can be used to predict the mean success rates of reasoners on related tasks. The natural logic approach highlights the inferential properties of expressions over their extensional ones and emphasizes the how we use natural language in inferences. While we show that NQL successfully models inferences on single and iterated quantifiers, we will also provide some empirically testable hypotheses that are derived from NQL's informative weights.

# Contents

# 1 Introduction

Logic and psychology can look back on a shared history that is full of twists and turns, with the two moving back and forth on their commitment to one another. Recently, van Benthem (2008) noted that logic is undergoing a cognitive turn, where only little earlier others proclaimed the time of logic in psychology to be over (see e.g. Evans 2002 or Chater & Oaksford 1999). And some time before that, Frege argued for the independence of logic and psychology with so much eager and success that van Benthem speaks of "Frege's wall" between the two fields. Logicians have since then however greatly extended the logical toolbox in some ways that can prove relevant for the investigation of a wide range of cognitive phenomena that we could describe as "reasoning". One of the corollaries of this extended toolbox is however that there is not *one* logic for this kind of task, but a variety of them (see for example Stenning & van Lambalgen (2010) and Besold *et al.* (2017) on how reasoning under uncertainty can be approached using logic and logic programming). Thus, when talking about logic in cognitive science and psychology, we do not restrict ourselves to what is usually implied by the umbrella-term "standard logic" (embodying predicate logic and propositional calculus) but a variety of very different tools that may or may not account for a variety of very different phenomena. We will see later on that much of the criticism by Evans (2002) and Chater & Oaksford (1999) can be traced back to such an equalization of logic with "standard logic". The story of "standard logic" in psychology is indeed analogue to the story of the hammer that mistakes everything for a nail and further fully disregards the pluralistic view of logic (e.g. Stenning & van Lambalgen 2012).

That means that we cannot just go about and "model reasoning with logic" but have to handpick a specific phenomenon of reasoning and the appropriate tool. The phenomenon of reasoning that caught our eye is that of inferences on quantifier expressions in natural language, e.g. ALL, NO, and MOST and their iterations, e.g. in "MOST pigeons annoyed AT LEAST THREE tourists".[1] Due to the limited scope of our investigation, we will restrict our discussion to specific quantifier expressions in the English language – determiners. While some researchers have emphasized the universality of quantification in noun phrases across all languages (Barwise & Cooper 1981, 177), we will later see that some languages do *not* use determiners for quantification. Quantifier expressions have been in logical spotlight since at least the advent of Aristotelian logic – with Aristotle being the (at least self-proclaimed) first logician and his syllogistic logic consequently the (at least self-proclaimed) first logic. There is thus much shared history to build upon. The appropriate logical hammer for this kind of reasoning, we think, is *natural logic*.

Natural logic emphasizes the fact that some important and recurring expressions in natural language are not only carriers of information, containers in which we store this and that piece of information, but allow for reasoning. Natural logic tries to make

[1] To avoid misunderstandings, we will consistently write quantifier expressions in small caps, e.g. MOST instead of "most".

sense of such expressions by giving them a formal characterization emphasizing the properties that allow for the reasoning in question – this operationalizes a variety of logical constants with semantic properties mirroring such in natural language expressions, e.g. ″and“, ″or“, ″necessarily“, ″all“ etc. Reasoning and inferences that build on the semantics of these expressions are thus inferences that operate ″on the surface of natural language“ (van Benthem 2007). As it happens, quantifier expressions show semantic properties that allow for a rich variety of such inferences.

We will present a natural logic – an extension of the model proposed by Geurts (2003) – that captures the essential inferential properties of single and iterated quantifier expressions. At the semantic center of this logic is the notion of monotonicity: we can infer that ″MOST pigeons annoyed AT LEAST THREE humans“ from ″MOST pigeons annoyed AT LEAST THREE tourists“ because, from an anthropocentric viewpoint, we know that the set of all tourists is contained in the set of all humans. This example already highlights the direction in which we will expand Geurts' logic – it contains two quantifier expressions and these *iterations* allow for monotonicity-based reasoning similar to syllogistic reasoning. We will introduce inference rules that account for this kind of reasoning as well. A natural logic defines some inferences as *good* and others implicitly as *bad* – but by adding a complexity measure to the inferences that are deemed good, we can also make statements about the *difficulty* of a good sequence of reasoning steps. We will create such a measure that predicts the mean success rate of participants in adequate reasoning-experiments concerned with (iterated) generalized quantifiers. The entirety of inference rules and weight assignments will carry the name natural quantifier logic (NQL).

This endeavor builds on the intersection of logic, the psychology of reasoning and linguistics and operates under the assumption that it is possible to find a measure of *computational complexity* of reasoning with quantifier expressions that aligns with the variation in its *cognitive difficulty* and correlates with mean success rates of human reasoners. Szymanik (2016) applies this idea to generalized quantifiers (where the complexity-measure is grounded in their representation as finite automata) but remains relatively silent on reasoning with them. The general approach of aligning complexity with difficulty is presented by Isaac *et al.* (2014). Due to the interdisciplinarity of the approach, there is no bottom-up way to build up a natural logic model of reasoning with (iterated) generalized quantifiers – we will take a rather circular approach and often introduce notions (e.g. that of monotonicity) by means of examples, before we can formally introduce them in a later chapter. We will thus circle back to the same ideas from different perspectives various times, hoping that they reciprocally clarify one another.

Firstly, in chapter 2, we will expand on the role of logic in the psychology of reasoning – where we focus on two natural language fragments that feature inferences with generalized quantifiers and iterated generalized quantifiers, respectively. As for the first fragment – Aristotle's syllogisms – we will see competing theories that try to explain human performance and make predictions regarding their difficulty. In evaluating these theories against each other, we will find our reasons to go with the natural logic approach, which we will introduce in some detail in chapter 3. This is mostly a a matter of *representation*: as natural logic uses natural language representations, it is both flexible and expressive enough to be cognitively plausible.

Secondly, we will introduce generalized quantifiers in chapter 4. We will focus on their relationship to natural language and their semantic properties that allow for the formulation of inferences and extend the relevant semantic notions to iterated quantifiers where possible. We will however also see that the interpretation of quantifier expressions such as MOST is not always clear – and that the inferential properties of generalized quantifiers sometimes depend on their precise interpretation. We will thereby motivate a simplicity constraint stating that interpretation-dependent inferences should only be included if they prove *useful*, and not just when they are *possible*.

Thirdly, we will present NQL, a natural logic for reasoning with (iterated) generalized quantifiers, in chapter 5 that is an extension of the model in Geurts (2003). NQL will account for a larger variety of single quantifiers, more inferences based on semantic properties, inferences on the combined semantic properties of iterated quantifiers and more. We will motivate the inference rules and their weight-assignments with a large array of empirical evidence and semantic considerations whenever possible. We see our main contribution in this extension of Geurts' model to a larger array of inferences, weight assignments that are better grounded in psychological research and thereby a complexity measure that better aligns with empirical data – we will see that the complexity measure of our logic will show a good correlation with mean success rates in experiments on syllogistic and related reasoning.

Finally, we will evaluate our logic in chapter 6 and see that it outperforms comparable models where comparison is possible and provides good predictions where not. We will further derive some testable predictions of our model in chapter 7, propose adequate experiments to test them and provide some philosophical context for our focus on the inferential properties of natural language expressions in chapter 8.

The natural logic approach brought forward here does not claim to give a full-fledged model of reasoning, we can rather state that a natural logic model for reasoning is almost necessarily incomplete, it presupposes a successful processing of interpretation and understanding (Braine 1978). It only claims that it is a very particular hammer that suits a very particular nail.

# 2 Of Logic and Reasoning

Variations of deductive reasoning have been heavily studied in psychology and logic throughout the last decades. But this paradigm and its logical import into psychology has come under pressure from various directions (e.g. Evans 2002, Chater & Oaksford 1999) – we will spend some time trying to defuse these criticisms and then argue in favor of a logic-based approach. After all, we owe the reader some motivation.

Critiques of investigations into deductive reasoning and the involvement of logic in these endeavors have often focused on the fact that "this paradigm was developed in a context of logicist thinking that is now outmoded“ (Evans 2002, 978) and that logic cannot be the proper normative standard against which human reasoning is to be evaluated. This is a reaction to somewhat logicist statements that "logical forms [...] are concerned with the ideal, with 'how we ought to think'.“ (Henle 1962, 366), thereby proclaiming logic the science of thought (ibid.). This more recent skepticism against logic and the deduction paradigm mirrors a change of ideas on rationality and thinking which comes with an increased focus on pragmatic and other non-logical factors in reasoning, e.g. propositional content (though we will see later on that a logic *can* account for pragmatic and other factors by assigning informative weights to inference rules). Throughout our work, we will focus on syllogistic and related modes of reasoning, i.e. a variety of reasoning forms closely related to what one could call *everyday deductions* like this one:

Some linguists are semanticists.<br>
All semanticists are philosophers.

---

Some linguists are philosophers.

Deductive reasoning however goes far beyond that: adopting Evans' (2002) terminology that later experiments by Wason and others (e.g. Wason 1983) introduced the *modern* paradigm, we might call syllogisms and their psychological investigation the *classical* paradigm. Kicking off the modern paradigm, Wason introduced the selection task to investigate the logical capacities of participants (this variant here is taken from Stenning & van Lambalgen 2012). For the experiment, a reasoner is confronted with four cards:

$\boxed{A}$ $\boxed{K}$ $\boxed{4}$ $\boxed{7}$

Reasoners are informed that they can only see one side (but not the other), and that each card has a number on one side and a letter on the other. Their task is to select the cards that one must turn to test the rule "If there is a vowel on one side, then there is an even number on the other side“ without turning any unnecessary cards. If propositional logic is taken to be the adequate normative standard, only 5% of all participants manage to select the correct cards (*A* and 7).

Evaluating the results against standard predicate logic, one can indeed only conclude that people are illogical and irrational (Evans 2002, 980). A way of rehabilitating logic was however found in the work of Stenning & van Lambalgen (2012), claiming that "the unargued adoption of classical logic as criterion of correct performance is thoroughly antilogical“ (Stenning & van Lambalgen 2012, 45) – and this for a variety of reasons. One can formalize the selection task in a multitude of logics which may even assign different interpretations to the logical constants: implication in standard logic is material (i.e. $p \rightarrow q$ is wrong if and only if $p$ and $\neg q$ are true), a standard that was unfortunately adopted by psychologists for their evaluation of how logical subjects reason in experimental setting. But the meaning of logical connectives is by no means fixed across all logics – a problem that refers to the correct *interpretation* of reasoning tasks (Evans 2012, 990). In this vein, Stenning & van Lambalgen (2012) explicitly distinguished between reasoning *to* and *from* an interpretation: a reasoner first needs an interpretation (what they call a logical parameter setting) before she can reason accordingly. The evaluation of Wason's selection task in terms of standard logic thus means evaluating subjects according to a normative standard that they did not apply themselves and were not in any way aware of. For the remainder of this work, we will however, for the most time, disregard reasoning *to* an interpretation and focus on reasoning *from* it: we will assume that the interpretations of logical constants are fixed as we introduce them further below. That being said, we will shortly see a natural logic with a directional entailment relation (Braine 1978) accounting for the results of the selection task.

While we do not wish to spend too much time on the general adequacy of logical modeling paradigms, it is probably safe to say that much of the recent criticism that logic had to endorse can be traced back to an uncalled-for equalization of logic and predicate logic or propositional calculus (with the umbrella-term *standard logic* usually being used). In Oxford & Chater's work, for example, "logic" is clearly to be read as standard logic, in effect equating the both. They write, for example, "...standard logic, which mental logic and mental models assume to be normative..." (Oaksford & Chater 2001, 349). [2] We will now look at the syllogistic fragment in some detail, compare theories that aim to explain experimental results and then go on to argue why logic is the right way to go here.

## 2.1 The Syllogistic Fragment

The analysis of syllogistic reasoning can look back upon a rich history – its logical investigation presumably started with Aristotle, while its psychological aspects came into focus some 100 years ago: Khemlani & Johnson-Laird (2012) note Störring (1908) as the first to investigate the psychology of deductive reasoning. Since then, the psychology of reasoning has seen a large variety of theories that aim to explain the why, the how, and the what of such inferential practices. We will take a moment to break this down here. According to a meta-study by Khemlani & Johnson-Laird (2012), there are three main kinds of theories regarding syllogistic reasoning: firstly, heuristic theories that emphasize principles that could underly *intuitive* responses. Secondly, two kinds of theories that emphasize *deliberative* reasoning – with methodological focus either on

[2] A logician can consequently respond to their favored example of logic being unable to account for non-monotonic inferences in everyday life by pointing out that there are indeed rich non-monotonic logics which can account for defeasible reasoning (Strasser & Antonelli 2016).

logical or set-theoretical notions. As a matter of further complication, these theories differ in what they call *good* and *bad* inferences.[3]

Syllogisms proof to be a good starting point for our endeavor, in fact, psychological studies that are concerned with human reasoning using quantification expressions are more often than not restricted to syllogistic reasoning. With the syllogistic fragment, we get a first study of quantification, a variety of inferential patterns and a large array of empirical data (assembled in a meta-study by Chater & Oaksford 1999, see table 1) that a model of syllogistic reasoning has to account for.

Syllogistic reasoning is deductive reasoning, and ″in daily life, individuals reason in a variety of contexts, and often so rapidly that they are unaware of having made an inference“ (Khemlani & Johnson-Laird 2012). Not all of this constitutes deductive reasoning and not all deductive reasoning is syllogistic reasoning – given, however, that humans are clearly able to perform deductive reasoning in everyday life (or at least while playing Sudoku or participating in a psychological experiment), this issue is worth investigating. Moreover, syllogistic reasoning has the advantage to be constituted of – formally – relatively simple reasoning patterns (some, however, prove to be – cognitively – quite difficult, as we will see) and empirical data of syllogistic reasoning can thus serve as a first benchmark for any formal theory of reasoning. By no means, however, is a theory of syllogistic reasoning to be confused with a theory of deductive reasoning and even less so with a theory of reasoning – as is exemplified by the selection task above.

Formally, syllogisms are inference schemes (inferences that hold for every proper assignment of variables) using the quantifier expressions EVERY, SOME, NO, and SOME…NOT. For historical reasons, we adopt the shortcuts ″A“ for EVERY, ″I“ for SOME, ″E“ for NO and ″O“ for SOME…NOT.[4] They consist of three quantified sentences (two premises and one conclusion) and three variables (which we will call $A$, $B$, and $C$). All three variables occur in the premises, while the one variable that occurs in both, called $B$, is not appearing in the conclusion. Aristotle noticed four *figures*, representing possible variable configurations in the premises. Figures do thus take different combinations of the variables $A$, $B$, and $C$ into account, but not yet the different assignments of quantifiers, which is called a syllogisms *mood*.

**Figures and Moods**
For $i \in \{1,2,3\}$, let $Q_i$ be any of the Aristotelian quantifiers. The four *figures*, representing the possible variable configurations are

[3] Throughout this investigation into different paradigms of syllogistic reasoning, we will see a variety of normative standards for reasoning, each of which makes sense in their own way. We will thus often not use the terminology of *validity* to classify a favored class of inferences but simply use *good*, which is less rich in implications and accounts for the relevance of non-logical theories.

[4] The symbols are derived from the first two vowels in *affirmativo* and *nego*, which mean ″I affirm“ and ″I deny“, respectively.

Figure 1 $$\begin{array}{c} Q_1(B,C) \\ Q_2(A,B) \\ \hline Q_3(A,C) \end{array}$$

Figure 2 $$\begin{array}{c} Q_1(C,B) \\ Q_2(A,B) \\ \hline Q_3(A,B) \end{array}$$

Figure 3 $$\begin{array}{c} Q_1(B,C) \\ Q_2(B,A) \\ \hline Q_3(A,C) \end{array}$$

Figure 4 $$\begin{array}{c} Q_1(C,B) \\ Q_2(B,A) \\ \hline Q_3(A,C) \end{array}$$

Assigning Aristotelian quantifiers to the $Q_i$ gives the syllogism its *mood*.

We will later see theories that use the concept of figures and moods to explain experimental results. Given that there are these 4 figures and 4 possible assignments for every $Q_i$ (thus $4^3$ moods), we end up with $4 \cdot 4^3 = 256$ possible syllogisms, most of which are *not* good inferences in any sense of the word. Lets look at a bad example to make the notation clear:

OA3E $$\begin{array}{l} \textsc{some}(B, \textsc{not } C) \\ \textsc{all}(B,A) \\ \hline \textsc{no}(A,C) \end{array}$$

Where 3 states the fact that the argument has figure 3, and O, A, and E state that the major premise is an O-proposition, the minor premise an A-proposition and the conclusion an E-proposition, respectively. Khemlani & Johnson-Laird (2012) however note, that there are 512 syllogisms, if one allows conclusions of the form $Q_3(C, A)$ (as it was done in scholastic logic). For the lack of a catchy name, we will henceforth call these 512 inferences the extended syllogistic fragment. Case in point is the Aristotelian syllogism AA4I, which is contrasted with the "extended" AA4A:

AA4I $$\begin{array}{l} \textsc{all}(C,B) \\ \textsc{all}(B,A) \\ \hline \textsc{some}(A,C) \end{array}$$

AA4A $$\begin{array}{l} \textsc{all}(C,B) \\ \textsc{all}(B,A) \\ \hline \textsc{all}(C,A) \end{array}$$

Note that the label alone does not tell us whether a conclusion is of the form $Q_3(C, A)$ or $Q_3(A, C)$. We will thus explicitly state if a label refers to a syllogism from the extended fragment. The Aristotelian AA4I with its restriction on the form of the conclusion is endorsed much less by participants in experimental settings than its extended counterpart AA4A – a fact that will later be mirrored in our natural logic, which even offers an explanatory account for such and similar inferential preferences.

Syllogisms lend themselves to psychological experimentation and to formalization: on the formal side, syllogisms are reasoning patterns in an apparently simple fragment of natural language. Furthermore, the subject-predicate form in syllogistic reasoning mirrors its presence in natural language. Syllogisms are thus of a form close to natural language and are readily extended to reasoning with other generalized quantifiers such as MOST, which are not expressible in first-order logic (see for example Chater & Oaksford 1999). On the psychological side, syllogisms can be presented in experimental settings as for the subjects to either make or judge a conclusion of given premises – a relatively simple experiment. Regarding this psychological perspective, Khemlani & Johnson-Laird (2012) however note the important fact that

> ″the principal moral of these results is that individuals use a variety of strategies in reasoning...“ (Khemlani & Johnson-Laird 2012, 6)

This carries within an important corollary: if reasoners use a wide variety of strategies, the modeling of a single strategy cannot be sufficiently explanatory (for a more extensive study of different strategies in syllogistic reasoning, see for example Newstead 1989).

## 2.2 Syllogistic Reasoning in Practice

The results of Chater & Oaksford's meta-study are presented in table 1. We will get a quick look at the multitude of explanatory accounts for their results before we zoom in on three theories that offer predictions regarding the cognitive difficulty of syllogisms.

One explanatory account is to focus on *figural effects*, i.e. that reasoners prefer certain conclusions in certain figures: subjects are biased towards A-C conclusions in figure 1, towards C-A conclusions in figure 2 and slightly biased towards A-C conclusions in figures 3 and 4, making conclusion against this bias difficult (Khemlani & Johnson-Laird 2012, 5). Note however that this presupposes the acceptance of A-C conclusions, which is contrary to the classical theory of syllogisms containing 256 syllogisms. The explanatory account of figural effect is thus only explanatory in the extended syllogistic fragment.

*Atmosphere* theory states that reasoners prefer to draw conclusions that fit the mood of the premisses and goes back to Sells (1936). More specifically, if one of the premises contains a NO, reasoners are biased towards a NO-conclusion as well and if one of the premises contains a SOME, reasoners are biased towards a SOME-conclusion. Otherwise, the bias is towards affirmative conclusions (Begg & Denny 1969). The takeaway is, that reasoners might take the quantifier expressions in the premises as hints towards which quantifier expression to use in the conclusion. Khemlani & Johnson-Laird (2012) note that most valid syllogisms indeed overlap with the atmosphere effect. *Figural effects* and *atmosphere* propose heuristics stating that variable assignment (figure) and quantifier assignment (mood) of the premises indicate the correct solution. They do however lead to conclusions when the correct answer would be that nothing follows from the premises and do not allow for statements about the cognitive difficulty of different tasks.

Other theories suggest that drawing conclusions from two syllogistic premises involves deliberate reasoning using sets. Using Venn diagrams, the three sets involved in syllogistic reasoning, $A$, $B$, and $C$, can be represented as three circles, a perspicuous

**Table 1:** Percentage of times a syllogistic conclusion was endorsed as reported in Chater & Oaksford (1999). All numbers are rounded to the closest integer and valid conclusions are marked as gray. If two conclusions are marked as valid, the first one is valid only in predicate logic. AA1A for example is a valid syllogism in predicate logic, while AA1I is in Aristotelian logic (relying on existential import).

| premises | conclusion | | | | premises | conclusion | | | | premises | conclusion | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| & figure | A | I | E | O | & figure | A | I | E | O | & figure | A | I | E | O |
| AA1 | 90 | 5 | 0 | 0 | AO1 | 1 | 6 | 1 | 57 | IO1 | 3 | 4 | 1 | 30 |
| AA2 | 58 | 8 | 1 | 1 | AO2 | 0 | 6 | 3 | 67 | IO2 | 1 | 5 | 4 | 37 |
| AA3 | 57 | 29 | 0 | 0 | AO3 | 0 | 10 | 0 | 66 | IO3 | 0 | 9 | 1 | 29 |
| AA4 | 75 | 16 | 1 | 1 | AO4 | 0 | 5 | 3 | 72 | IO4 | 0 | 5 | 1 | 44 |
| AI1 | 0 | 92 | 3 | 3 | OA1 | 0 | 3 | 3 | 68 | OI1 | 4 | 6 | 0 | 35 |
| AI2 | 0 | 57 | 3 | 11 | OA2 | 0 | 11 | 5 | 56 | OI2 | 0 | 8 | 3 | 35 |
| AI3 | 1 | 89 | 1 | 3 | OA3 | 0 | 15 | 3 | 69 | OI3 | 1 | 9 | 1 | 31 |
| AI4 | 0 | 71 | 0 | 1 | OA4 | 1 | 3 | 6 | 27 | OI4 | 3 | 8 | 2 | 29 |
| IA1 | 0 | 72 | 0 | 6 | II1 | 0 | 41 | 3 | 4 | EE1 | 0 | 1 | 34 | 1 |
| IA2 | 13 | 49 | 3 | 12 | II2 | 1 | 42 | 3 | 3 | EE2 | 3 | 3 | 14 | 3 |
| IA3 | 3 | 85 | 1 | 4 | II3 | 0 | 24 | 3 | 1 | EE3 | 0 | 0 | 18 | 3 |
| IA4 | 0 | 91 | 1 | 1 | II4 | 0 | 42 | 0 | 1 | EE4 | 0 | 3 | 31 | 1 |
| AE1 | 0 | 3 | 59 | 6 | IE1 | 1 | 1 | 22 | 16 | EO1 | 1 | 8 | 8 | 23 |
| AE2 | 0 | 0 | 88 | 1 | IE2 | 0 | 0 | 39 | 30 | EO2 | 0 | 13 | 7 | 11 |
| AE3 | 0 | 1 | 61 | 13 | IE3 | 0 | 1 | 30 | 33 | EO3 | 0 | 0 | 9 | 28 |
| AE4 | 0 | 3 | 87 | 2 | IE4 | 0 | 1 | 28 | 44 | EO4 | 0 | 5 | 8 | 12 |
| EA1 | 0 | 1 | 87 | 3 | EI1 | 0 | 5 | 15 | 66 | OE1 | 1 | 0 | 14 | 5 |
| EA2 | 0 | 0 | 89 | 3 | EI2 | 1 | 1 | 21 | 52 | OE2 | 0 | 8 | 11 | 16 |
| EA3 | 0 | 0 | 64 | 22 | EI3 | 0 | 6 | 15 | 48 | OE3 | 0 | 5 | 12 | 18 |
| EA4 | 1 | 3 | 61 | 8 | EI4 | 0 | 2 | 32 | 27 | OE4 | 0 | 19 | 9 | 14 |
| | | | | | | | | | | OO1 | 1 | 8 | 1 | 22 |
| | | | | | | | | | | OO2 | 0 | 16 | 5 | 10 |
| | | | | | | | | | | OO3 | 1 | 6 | 0 | 15 |
| | | | | | | | | | | OO4 | 1 | 4 | 1 | 25 |

| | |
|---|---|
| A = all | E = no |
| I = some | O = some...not |

representation that allows for the making of inferences (e.g. Shin 1992). This yields a feasible method that is at the same time sufficiently flexible – Peirce's diagrams for example can account for all of predicate calculus but, according to their creator, lose all psychological plausibility (Peirce 1958). Furthermore, number of researchers have considered *natural set theory*, a hypothesis stating that humans have a natural way to deal with a plurality of objects (e.g. Seuren 2010). While this hypothesis has, to the best of our knowledge, not been tested yet, this could yield an alternative to create formal models of syllogistic reasoning focusing on sets.

Finally, we will ignore the complicating issue that there are large individual differences in syllogistic reasoning: Khemlani & Johnson-Laird (2012) report that in experiments at a highly selective university, 55% of all inferences made were valid, whereas in experiments at a non-selective university, 37% of all inferences made were valid. Further, the proportion of valid inferences drawn per participant ranges from 85% to 15% validity. While these differences are important to keep in mind, we will henceforth work with the mean success rates reported in the meta-study by Chater & Oaksford (1999) as in table 1. Figural effects, atmosphere theory and Venn diagrams have in common that they suggest solutions to syllogisms but not their difficulty – we will now look at three paradigmatic theories that have something to say about that.

## 2.3 Formal Accounts of Syllogistic Reasoning

An immediate observation is, unsurprisingly, that some *good* inferences are easier than others. We will henceforth refer to this as the *cognitive difficulty* of an inference – the three formal accounts below differ from atmosphere, figural effects and Venn diagrams in the sense that they not only want to give the correct conclusion (or the one endorsed by most reasoners) but also a measure of complexity that aligns with the cognitive difficulty of inferences. The guiding principle is thus to develop a measure of *computational complexity* that aligns with, or even predicts, the *cognitive difficulty* observed in experiments. As a matter of further complication, all these theories yield different normative standards, i.e. labeling different syllogistic inferences as *good*.

### 2.3.1 Natural Logic

An early account that aimed to investigate the psychology of reasoning with a logical model was given by Rips (1983 and 1994). Reasoning was thereby understood as proof in a natural deduction system while the cognitive difficulty of a sequence of inferences was conceptualized as the length of proof of a conclusion from assumptions. Rips introduced further inference rules to account for the fact that some inferences, e.g. the syllogism AA1A, are of low cognitive difficulty while its proof in natural deduction systems takes seven steps. While Rips' work is certainly to be understood as the ancestor of much work into natural logics that followed, it became increasingly clear that it faces a variety of problems connected to its first order logic representations (exhaustively discussed in

Johnson-Laird 1997). We will thus focus on a later model brought forward by Geurts (2003), which we will also extend later on. We will however later revisit the distinction between natural deduction and natural logic and their common focus on the inferential properties of logical constants that correspond to natural language expressions. The general idea of focusing on monotonicity-properties of quantifiers as to create a natural logic for reasoning with them stems from van Benthem (1986, Chapter 7) and Sanchez Valencia (1991).

Exemplifying (natural) logic based approaches to syllogistic reasoning with the work of Geurts (2003) brings with it the benefit of already getting a first glimpse at what will be happening later. They present a natural logic that emphasizes the importance of monotonicity for comprehension of quantifiers and inferences involving quantification. Their logic pivots on semantical properties of quantifiers and aims to show that the study of formal semantics – here in the case of generalized quantifiers – provides insight for the psychology of deduction. Similar to natural deduction systems, inference rules lie at the heart of the natural logic approach, highlighting the close tie between inference and interpretation. Monotonicity will be defined formally below but we can already introduce the idea by means of examples here. Consider the sentences

(i) All flowers are vermilion. ($\textsc{all}(A, B)$)

(ii) No flowers are red. ($\textsc{no}(A, C)$)

Sentence (i) entails ”All flowers are red“ ($\textsc{all}(A, C)$) while (ii) entails ”No flowers are vermilion“ ($\textsc{no}(A, B)$) because the set of all vermilion things is a subset of all red things ($\textsc{all}(B, C)$). More specifically, we have a variant of the syllogism AA1A:

$$\text{AA1A} \qquad \frac{\begin{array}{c}\textsc{all}(\text{vermilion, red})\\ \textsc{all}(\text{flowers, vermilion})\end{array}}{\textsc{all}(\text{flowers, red})}$$

We will usually say that the determiners ALL and NO put their second argument in upward- and downward-entailing positions, respectively (are right-side upward monotone and right-side downward monotone, respectively). We can already see how this allows for monotonicity-based inferences as the ones in (i) and (ii) above:

$$\text{Mon}\uparrow \qquad \frac{\begin{array}{c}Q\uparrow(A,B)\\ \textsc{all}(B,C)\end{array}}{Q\uparrow(A,C)} \qquad\qquad \text{Mon}\downarrow \qquad \frac{\begin{array}{c}Q\downarrow(A,B)\\ \textsc{all}(C,B)\end{array}}{Q\downarrow(A,C)}$$

where Q $\downarrow$ $(A, B)$ and Q $\uparrow$ $(A, B)$ mean that the quantifiers put their second argument in a downward and upward entailing position, respectively.[5] Geurts' logic is part of a tradition stating that syllogistic reasoning (and by extension reasoning with

[5] We wish to add that we use different names as well as different notation than Geurts (2003) – this is due to the fact that these rules will be expanded and then referred back to throughout our work.

generalized quantifiers) is essentially monotonicity-based reasoning. While these two inference rules do much of the work required in the syllogistic fragment, some other rules are necessary to account for all of it:

$$\text{Conv} \qquad \frac{Q(A,B)}{Q(B,A)} \qquad\qquad \text{pConv} \qquad \frac{\textsc{no}(A,B)}{\textsc{all}(A,\textsc{not}\ B)}$$

where Conv (symmetry) is only applicable if $Q$ is either SOME or NO and hinges on the symmetry property that we will introduce formally later and pConv (pseudoconversion) only works with NO. These rules are already sufficient to prove all 15 syllogisms that are valid in predicate logic. Foreshadowing our later discussion on existential import, we need another rule to make all syllogisms that are valid in Aristotelian logic provable (Geurts 2003, 243):[6]

$$\text{exImp} \qquad \frac{\textsc{all}(A,B)}{\textsc{some}(A,B)}$$

As to fit the empirical data provided by Chater & Oaksford (1999), Geurts (2003) assign a *cost* to each inference rule. A reasoner starts out with 100 points of which the costs of the inferences are subtracted: each use of either Mon↑ or Mon↓ costs 20 points, every use of pConv 10 points and every proof that contains an O-proposition (particular negative) costs an extra 10 points (thus implying that Conv and exImp are for free). Geurts' model leads to the predictions in table 2.

**Table 2:** Comparison between the experimental data in Chater & Oaksford's (1999) meta-study (in brackets) and Geurts' model. Syllogisms are ordered in decreasing order of mean success. Predictions that are more than 10% apart from the actual performance are marked as gray.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| AI1I | (92) | 80 | IA3I | (85) | 80 | EA3O | (22) | 40 |
| IA4I | (91) | 80 | OA3O | (69) | 70 | AA4I | (16) | 60 |
| AA1A | (90) | 80 | AO2O | (67) | 70 | EA4O | (8) | 40 |
| AI3I | (89) | 80 | EI1O | (66) | 60 | AA1I | (5) | 60 |
| EA2E | (89) | 80 | EI2O | (52) | 60 | EA1O | (3) | 40 |
| AE2E | (88) | 80 | EI3O | (48) | 60 | EA2O | (3) | 40 |
| EA1E | (87) | 80 | AA3I | (29) | 60 | AE4O | (2) | 40 |
| AE4E | (87) | 80 | EI4O | (27) | 60 | AE2O | (1) | 40 |

Their model reaches a good fit with the syllogisms that constitute *good* sequences of inferences from an Aristotelian or predicate logic viewpoint: the model predicts much of the variance that occurs in mean success ($r^2 = 0.87$) and its predictions show strong

[6] This system allows a solution to more than just the syllogisms that are valid in Aristotelian or predicate logic. Some of the additional ones are proven in appendix B.

correlation with actual human performance (Pearson r $= 0.93$).[7] Zhai *et al.* (2015) used the same logic but had the weights directly learned from the data – we do however think that the ratio of parameters to datapoints invites overfitting, eventually making the weights less informative. We wish to point out two weak spots of Geurts' theory: the theory only applies to valid syllogisms, highlighting a lack of generality. Furthermore, the weights of the inference rules seem somewhat arbitrary. Both critiques were already pointed out by the authors themselves.

### 2.3.2 Mental Models

Mental models theory has its foundation in the work of Johnson-Laird (e.g. Johnson-Laird & Bara 1984, Johnson-Laird 2010). Its assumption is that reasoners construct mental models which are consistent with the information they have received so far, where each mental model represents a possibility that is consistent with the present information state. So if $A$ and $B$ are two atomic propositions, the mental models for statements using operators from standard logic would be as in table 3.

**Table 3:** Fully explicit mental models representations for some logical connectives. "*A* and *B*" for example is represented by one model, others by two or three.

| Connective | MMs | |
|---|---|---|
| $A$ and $B$ | $A$ | $B$ |
| $A$ xor $B$ | $A$ | $\neg B$ |
| | $\neg A$ | $B$ |
| $A$ or $B$ | $A$ | $B$ |
| | $A$ | $\neg B$ |
| | $\neg A$ | $B$ |
| If $A$ then $B$ | $A$ | $B$ |
| | $\neg A$ | $B$ |
| | $\neg A$ | $\neg B$ |
| $A$ iff $B$ | $A$ | $B$ |
| | $\neg A$ | $\neg B$ |

We can already see that these connectives differ in how many mental models they require for their representation – which can be used to measure their difficulty (Johnson-Laird 2010, 3). As every model (one line) is a possibility, a conclusion is deemed necessary if it holds in all of the models. $A$ is thus a necessary conclusion from the premise "*A* and *B*" but not from the premise "*A* xor *B*". Mental model theory assumes that reasoners represent sets by building mental models of their members, aim to maintain semantic information, are parsimonious and engage in a search for counterexamples against a

[7] Geurts (2003) reported $r = 0.83$ – the analysis presented here is made on the data in table 2.

model (Khemlani & Jonson-Laird 2012, 19). A mental model of ”All semanticists are philosophers“ would thus look like the following:

| | |
|---|---|
| semanticist | philosopher |
| semanticist | philosopher |
| semanticist | philosopher |
| | philosopher |
| | philosopher |
| | ... |

This highlights that the set of semanticists might not fully exhaust the set of philosophers. According to Johnson-Laird (2010, 2), this kind of representation is *iconic*, i.e. corresponds to what it represents as much as possible (this is opposed to a *symbolic* representation, whose form has only conventional connections to its content). As is custom in mental model theory, we will take a step away from this iconic representation and adopt the convention that squared brackets around a letter mean a set is exhaustively represented by a symbol. More concrete, we will understand ”All semanticists are philosophers“ as

| | |
|---|---|
| [semanticist] | philosopher |
| | ... |

instead of a representation akin to the one above, with [...] being an implicit model for all cases where the antecedent is wrong. Let us apply this to syllogisms (examples from Khemlani & Johnson-Laird 2012). On the left hand side, we see two syllogistic premises of the form AE4, on the right side its representation as a mental model.

| | | | |
|---|---|---|---|
| All A are B. | [A] | [B] | ¬C |
| No B are C. | [A] | [B] | ¬C |
| | | | [C] |
| | | | [C] |
| | | ... | |

This conjoins mental models for both premises: all *A*s are *B*s, as noted above, is represented by

| | |
|---|---|
| [A] | B |
| | ... |

whereas no *B*s are *C*s by

| | |
|---|---|
| [B] | ¬C |
| | [C] |
| | ... |

stating that everything that is a *B* is a not-*C* and that there are other things which are *C*s. Note that in syllogistic reasoning, only the relationship between *A* and *C* is of interest. Combined, this states that all *A*s are not-*C*s , that there are *C*s and only things that are not-*A*s can be *C*s, yielding the conclusion ″No *A* are *C*“. We will now see a pair of premises (AI4) that requires two mental models:

| | | | | | | |
|---|---|---|---|---|---|---|
| All A are B. | [A] | B | C | [A] | B | |
| Some B are C. | [A] | B | | | B | C |
| | | | C | [A] | B | |
| | | ... | | | | C |
| | | | | | ... | |

While the first model yields the conclusion ″Some *A* are *C*“, the second model is a counterexample to this conclusion – the correct answer is thus that nothing follows from premises AI4: no statement about the relation between *A* and *C* is true in all models consistent with the premises. Mental model theory gives raise to a complexity measure, that is somewhat related to that of proof-length which will be introduced further below: the more models a syllogism takes, the more difficult it is.

This search for counterexamples – akin to falsificationism in philosophy of science – that the mental model theory imposes upon reasoners is however not supported by empirical evidence: reasoners do not try to engage in falsification but verification (Khemlani & Johnson-Laird 2012, 19). Furthermore, mental model theory suffers from a lack of flexibility – it is not clear how quantifiers beyond the syllogistic ones can be expressed in mental model theory (Chater & Oaksford 1999).

#### 2.3.3 Probability Heuristics

Theories of reasoning that invoke logical tools and methodology have recently been under pressure from theories that try to give reasoning – even the presumably paradigmatic case of the usefulness of logic, deductive reasoning – a probabilistic foundation. The probability heuristic model has been proposed by Chater & Oaksford (1999) and Oaksford & Chater (2001). They state that human reasoning in experimental settings is usually evaluated against the wrong normative standards – the right normative standard being probability heuristics, not logic. Thus, if human performance is evaluated against the standard of logic, the results of reasoning experiments shed bad light on human capacities, as they tend to make logically bad inferences. If one would however evaluate human performance against probability heuristics, there would be another story to tell: then, humans would carry their everyday-reasoning heuristics into the experimental setting and reason according to them. The evaluation would then have to conclude, that humans manage to follow their normative standard, given by probability heuristics, even in an experimental setting (Oaksford & Chater 2001, 349).

As for their model, it is relying on five heuristics, three of them generate conclusions (G1-G3) and two of them test them (T1-T2):

- **G1:** choose the quantifier of the conclusion to be the same as the quantifier in the least informative premise (the *min-premise*) where informativeness is an (intuitive) ordering of the propositions involved in syllogistic reasoning (A>I>E>O) that we will come back to in chapter 5.
- **G2:** the next most preferred conclusion is the one that is probabilistically entailed by the one from G1 (the *min-conclusion*).
- **G3:** if just one of the possible conclusion subject noun phrases matches the subject noun phrase of just one premise, then the conclusion has that subject noun phrase.
- **T1:** be confident in the conclusion generated by G1-G3 in proportion to the informativeness of the most informative premise (the *max-premise*).
- **T2:** avoid producing or accepting SOME…NOT conclusions.

They furthermore introduce what they call *probabilistic semantics* for quantifiers, thus defining quantifiers by means of conditional probabilities (Chater & Oaksford 1999, 200). So, for example, $\textsc{all}(A, B)$ means that the conditional probability of $B$, given $A$, is 1, i.e. $P(B|A) = 1$ and $\textsc{some}(A, B)$ means that $P(B|A) > 0$ and that there are things that are both $A$ and $B$ (existential import). While this approach makes it easy to extend their model beyond the syllogistic fragment, it is not clear how the probabilistic semantics of cardinal quantifiers or even iterated quantifiers can be given. Let us look at two examples of syllogistic reasoning using probability heuristics (taken from Chater & Oaksford 1999).

$$\text{AI1I} \qquad \begin{array}{ll} \textsc{all}(Y, X) & \textit{max-premise} \\ \textsc{some}(Z, Y) & \textit{min-premise} \\ \hline \text{I-type conclusion} & \textit{by G1} \\ \textsc{some}(Z, X) & \textit{by G3} \end{array}$$

$$\text{IE2E} \qquad \begin{array}{ll} \textsc{some}(X, Y) & \textit{max-premise} \\ \textsc{no}(Z, Y) & \textit{min-premise} \\ \hline \text{O-type conclusion} & \textit{by G2} \\ \textsc{some}(X, \textsc{not}\ Z) & \textit{by G3} \\ \text{E-type conclusion} & \textit{by G1} \\ \textsc{no}(Z, X) & \textit{by G3} \end{array}$$

Where the latter is valid in neither predicate logic nor Aristotelian logic. Oaksford and Chater give an interesting argument for favoring probabilistic over logical models that directly relates to our cause. They state that

> ”The most important feature of PHM [probability heuristic model] is that it can generalize to syllogisms containing quantifiers, such as *Most* and *Few*, that have no logical interpretation“ (Oaksford & Chater 2001, 354)

As we will see later on, this again is a claim that can only be supported by equating logic and predicate calculus. As we promote deductive inferences on natural language surface, we can rely on the expressiveness of natural language and interpretations of quantifiers that rely on quantification over sets, not only over individuals. Generality of the kind that is described by Oaksford and Chater is even a strength of the natural logic approach presented later.

A proper reason for the appreciation of the probability heuristic model might thus be to aim for a unifying theory of reasoning, i.e. the aim to formalize different cognitive capacities in one theory, e.g. probability theory. However, we wish to point out that probabilistic models, while sailing under the unifying flag of *probabilistic modeling* still have large variation, like different logics. A proper comparison of the two approaches would thus mean to compare all probabilistic models with all logical models.

### 2.3.4 Evaluation

We can now compare the three main approaches introduced above – some results can be seen in table 4.[8] No matter what preference one has, these results are at least a little bit funny: all three theories distinguish three identical categories of cognitive difficulty on this set of syllogisms (indicated by the additional horizontal lines in the table). If one was to choose their favorite theory of syllogisms, this choice can thus not be made on grounds of fit with empirical data and predictions. Note furthermore that all three theories, by making the same predictions, classify IE1O and EI1O as equally hard – while the mean of correct responses is 16% in the former and 66% in the latter. This calls for a more flexible complexity-measure that allows for differentiating between IE1O and EI1O in terms of cognitive difficulty.

All three theories have shown advantages and disadvantages. The reason to extend Geurts' (2003) logic is twofold: firstly, it offers generality in the sense that it can account for a larger variety of natural language quantifier expressions (as it uses natural language representations, it inherits parts of the flexibility of natural language) and is thus easily extended to quantifiers such as MOST while also carrying with it the normative aspect that some quantifiers such as FEW do not allow for many inferences because they have unfavorable monotonicity properties. This is contrasted by mental model theory's inability to go beyond Aristotelian quantifiers and the fact that the PHM's probabilistic semantics cannot account for cardinal quantifiers. Its generality continues in the sense that we can apply an extended version of Geurts' logic to iterated quantifiers that we will introduce later on – it is hard to see how probability heuristics or mental models could operationalize iterated quantifiers while natural logic can rely on their linguistic treatment in generalized quantifier theory and monotonicity-entailments similar to the ones introduced above. We thus plan to go beyond the syllogistic fragment in ways that are supported by neither mental model theory or probability heuristics. Secondly, their

[8] A small disclaimer: some of the numbers reported in Oaksford & Chater (2001) differ from those in the original study in Chater & Oaksford (1999) – the success rates for the EI- and IE-syllogisms have been switched, e.g. the success rate for EI3O has been switched with that for IE3O. In such cases, we will stick to the original numbers reported in Chater & Oaksford (1999).

**Table 4:** Comparison of mental model theory (MM), the probability heuristic model (PHM) and Geurts' natural logic (NatLog) for some valid syllogisms. For mental model theory, we count the number of mental models involved in solving the syllogism, for PHM we state the heuristics involved (both taken from Oaksford & Chater 2001) and for NatLog the scores in Geurts' system. The entries in the NatLog-column labeled with an asterisk were not provided by Geurts (2003) themselves and the one entry without a number is not a valid inference in Geurts' system. The additional proofs in Geurts' model are in appendix B.

| Syllogism | Theory | | | Mean (%) |
|---|---|---|---|---|
| | MM | PHM | NatLog | |
| AA1A | 1 | G1 | 80 | 90 |
| AI1I | 1 | G1 | 80 | 87 |
| IA3I | 1 | G1 | 80 | 88 |
| AI3I | 1 | G1 | 80 | 89 |
| IA4I | 1 | G1 | 80 | 88 |
| EA1E | 1 | G1 | 80 | 92 |
| AE2E | 1 | G1 | 80 | 85 |
| EA2E | 1 | G1 | 80 | 89 |
| AE4E | 1 | G1 | 80 | 91 |
| AO2O | 2 | G1+T1 | 70 | 67 |
| OA2O | 2 | G1+T1 | -* | 56 |
| AO3O | 2 | G1+T1 | 70* | 66 |
| OA3O | 2 | G1+T1 | 70 | 69 |
| EI1O | 3 | G2+T2 | 60 | 66 |
| IE1O | 3 | G2+T2 | 60* | 16 |
| EI2O | 3 | G2+T2 | 60 | 55 |
| IE2O | 3 | G2+T2 | 60* | 30 |
| EI3O | 3 | G2+T2 | 60 | 48 |
| IE3O | 3 | G2+T2 | 60* | 33 |
| EI4O | 3 | G2+T2 | 60 | 27 |
| IE4O | 3 | G2+T2 | 60* | 44 |

complexity measure proofs to be the most flexible – as the inference rules are weighted, derivations relying on different rules will yield a different measure of complexity. By this, we hope to obtain a better fit to the data and some testable predictions – a goal that we think both mental models and probability heuristics have too a static complexity measure for. Predictions of this kind are possible as weights assigned to inference rules are *informative*, i.e. have a concrete interpretation.

### 2.4 Beyond Syllogisms: Iterating Quantifiers

Syllogisms is not all there is to reasoning. Somewhat surprisingly, not much empirical investigations have been done that extend the existing data on syllogistic reasoning in a direction that is of interest for one who studies the psychology of quantification. Geurts & van der Silk (2005) did an experiment on reasoning with iterated quantifiers investigating how their combined monotonicity properties interact with the cognitive difficulty of inferences. For lack of a better name, we will henceforth call the natural language fragment involved in their experiment on reasoning the quantifier iteration (QI) fragment. We have already seen examples of monotonicity-based inferences using single quantifiers – the inferential properties of iterated quantifiers will turn out to be quite similar. Consider the following inference:

$$\begin{array}{l}\text{Most pigeons annoyed more than three tourists.}\\ \text{All tourists are human.}\\ \hline \text{Most pigeons annoyed more than three humans.}\end{array}$$

Understanding a formula $\phi(A, B)$ to mean that ″$A$ annoyed $B$“, we can formalize this as

$$\begin{array}{l}Q_1Q_2\phi(A,B)\\ \textsc{all}(B,C)\\ \hline Q_1Q_2\phi(A,C)\end{array}$$

where $Q_1$ is MOST and $Q_2$ is MORE THAN THREE. This inference is a relatively easy one – the interaction of the two quantifiers involved is just of the right kind: both MOST and AT LEAST THREE are right-side upward monotone, and put the argument $B$ in an upward-entailing position.

Other combinations are harder: Geurts & van der Silk (2005) note that the monotonicity properties of quantifiers have clear combinatorial properties which allow for a straightforward generalization of monotonicity onto iterated quantifiers (as exemplified in the example above), we will later see some counterexamples to this assertion. For the fragment that they introduce, we can however observe some regularities: when both $Q_1$ and $Q_2$ have the same right-side entailment properties, they together put the second argument in an upward entailing position, if they have different right-side entailment properties, they together put the second argument in a downward entailing position. Geurts & van der Silk (2005) note the following in their evaluation: inferences that involve upward entailment are easier than inferences that involve downward entailment, inferences that involve two quantifiers with the same (right-side) monotonicity properties are easier than those who do not and inferences that involve downward entailing cardinal quantification (e.g. using AT MOST FIVE) involve more cognitive difficulty. Participants in their study had to determine whether reasoning patterns of the form

**Table 5:** Success rates in the QI-fragment, where the numbers in the %-row represent the percentage of successful assessments of an argument as valid or not. Mean response for inferences that conform to the given definition of validity are marked as gray.

| $Q_A$ | $Q_B$ | Minor | % | $Q_A$ | $Q_B$ | Minor | % |
|---|---|---|---|---|---|---|---|
| EVERY | MORE THAN | ALL$(B, C)$ | 91 | MOST | MORE THAN | ALL$(B, C)$ | 91 |
| | | ALL$(C, B)$ | 69 | | | ALL$(C, B)$ | 67 |
| | FEWER THAN | ALL$(B, C)$ | 71 | | FEWER THAN | ALL$(B, C)$ | 62 |
| | | ALL$(C, B)$ | 58 | | | ALL$(C, B)$ | 60 |
| AT LEAST | MORE THAN | ALL$(B, C)$ | 96 | SOME | MORE THAN | ALL$(B, C)$ | 87 |
| | | ALL$(C, B)$ | 69 | | | ALL$(C, B)$ | 67 |
| | FEWER THAN | ALL$(B, C)$ | 53 | | FEWER THAN | ALL$(B, C)$ | 60 |
| | | ALL$(C, B)$ | 51 | | | ALL$(C, B)$ | 62 |
| AT MOST | MORE THAN | ALL$(B, C)$ | 51 | NO | MORE THAN | ALL$(B, C)$ | 69 |
| | | ALL$(C, B)$ | 38 | | | ALL$(C, B)$ | 53 |
| | FEWER THAN | ALL$(B, C)$ | 36 | | FEWER THAN | ALL$(B, C)$ | 73 |
| | | ALL$(C, B)$ | 49 | | | ALL$(C, B)$ | 64 |

$$\begin{array}{l} \text{Q}_A \text{ A played against Q}_B \text{ B.} \\ \text{All B were C. / All C were B.} \\ \hline \text{Q}_A \text{ A played against Q}_B \text{ C.} \end{array}$$

were valid or not with $Q_A \in \{$EVERY, MOST, AT LEAST, SOME, AT MOST, NO$\}$ and $Q_B \in \{$MORE THAN, FEWER THAN$\}$ and only one of the two possibilities of the minor premise present. That means, essentially, participants had to decide whether a presented argument used a valid monotonicity inference. The results can be seen in table 5. Participants got a definition of validity of the form "*If* the premises are true, the conclusion *must* be true as well.“. There are several things to note about this fragment: firstly, other than syllogisms, it is exclusively concerned with monotonicity-inferences. Secondly, those inferences are concerned only with the second arguments of quantifier expressions. We will see later on that Geurts' (2003) logic can readily be extended to account for monotonicity-inferences on iterated quantifiers in the QI-fragment and beyond. We will however also meet the limits of the (combinatorial) monotonicity-grounded approach as some quantifier expressions, e.g. MOST, have no left-side monotonicity properties and not all monotonicity properties combine in the same way.

# 3 Natural Logic – A Logical Approach to Reasoning

We have seen in both datasets that some inferences prove to be easy, while others are hard an show success-rates below chance-level. Consequently, Braine (1990, 133) notes that, for many people, ″logical reasoning seems to present an odd and difficult, but erudite, mixture of the obvious and the counterintuitive“. We are convinced that much of this variation in cognitive difficulty can be explained by a semantic analysis of the expressions involved. The natural logic approach, in short, tries to present a logic whose inferential properties capture essential syntactic or semantic properties of the natural language fragment it aims to model. We will now talk about natural logic in general and then converge to a closer look at some natural logics. The first step in the direction of natural logic models is to look at natural deduction.

While natural deduction is related to other systems of deduction, e.g. Hilbert-Frege style proof systems, it does have a different emphasis: the goal is to create a system focusing on *inferences*, where said inferences are specified in a cognitively more plausible way or are at least not too cumbersome. The meaning of the logical constants in natural deduction (connectives and quantifiers) is defined by introduction and elimination rules – but still the same as in standard logic. These rules specify when and how one can make inferences that feature the operator in question as main operator in the conclusion (introduction rules) or the premises (elimination rules).

**Conjunction-Introduction and -Elimination**
Introduction and elimination rules state when one can draw an inference that introduces or eliminates a logical connective. See here introduction (I) and elimination (E) rules for conjunction:

(I) $$\frac{A_1 \qquad A_2}{A_1 \wedge A_2}$$ (E) $$\frac{A_1 \wedge A_2}{A_i}$$

where $i \in \{1, 2\}$.

We will omit the rest of Gentzen's system as we are only travelling through and its rules are overly familiar to logicians anyway. An inference rule validates the transition from one step to the next in an argument and this approach reflects the opinion that inference rules are cognitively more plausible than rules of proof and axioms. Whereas in a Hilbert-Frege style system, every inference has to be brought back to the axioms, this is not the case in natural deduction (see Sundholm 1983). But a system building up on axioms not only makes claims about human reasoning, but also its foundations and primitives and is largely inadequate as a model of human reasoning (see Braine 1978, 3 for further discussion of this point). We will return to the idea of favoring the inferential properties of logical constants that aim to mirror that of certain natural expressions (like Gentzen's rules for ″$\wedge$“ mirror the inferential properties of the natural language expression ″and“) at a later point. For now we should just realize that this is similar to our natural logic approach: both highlight the semantic properties of logical constants that lead to inferences.

Gentzen's theorem shows us how expressive his natural deduction system is – it clearly proves powerful (in fact equivalent to Hilber-Frege style proof systems), but still cognitively inadequate and limited in expressive power: most generalized quantifiers lie beyond its expressibility. Equally important is, however, that with language, we do not only have a means of saying something or expressing some information, but there are also forms of reasoning that operate directly on "the surface of natural language" (van Benthem 2007, 5). A natural logic is not only interested in inferences but in inferences that relate strongly to natural language. We thus have two distinct justifications for our pick of natural logic: other systems of inference are not expressive enough and we wish to have inferences operating on natural language representations (as we will see in more detail later, $\textsc{some}(A, B)$ is just a much more plausible representation than $\exists x(A(x) \wedge B(x))$ – Barwise and Cooper call this the "norotious mismatch" between the representational capacities of predicate calculus and natural language). Natural logic is thus not only about reasoning, but also about language and how the two interact: some parts of natural language allow for the extraction of formal properties that in turn allow for the modeling of reasoning.[9]

We will have a further discussion on natural logics and how to situate them in a cognitive modeling enterprise later on in chapter 8. For now, we will take a quick look at a variety of natural logic models to get a get a glimpse of what they might do.

## 3.1 Natural Logic – the Road so Far

Despite these introductory remarks, it is still not easy to pin down what natural logic actually is – a natural logic is a set of inference rules that act on the surface of natural language. But that does not come with any a priori constraints on how a natural logic should look like. We can however say that with each natural language, we get a variety of terms that allow for reasoning, e.g. expressions such as "or", "and", and "necessarily", but most importantly – in our case – quantifier expressions such as determiners. These natural language expressions allow their users to reason with them. We have already seen in the last chapter that such a set of inference rules for simple natural language reasoning can offer a simple, but informative model of reasoning that aligns reasonably well with human performance. This already outlined the path that we wish to travel but before we go further into this direction, we will have a quick stop and look at what is left and right of us. We will introduce various formal approaches to the capturing of inferential properties of natural language expressions and then decide why we can or

[9] There is no context-independent criterion whether more or less expressive power is better. The logic used needs to be primarily *adequate*. This concerns the natural language fragment that they aim to provide insight about (more often than not: Aristotle's syllogisms) and the goal of the modelers: some have rather cognitive motivations (emphasizing the reproduction of empirical data) while others have rather logical motivations (thus emphasizing metalogical properties such as completeness and soundness).

cannot call them a natural logic.

Braine (1978) introduces a natural propositional logic that tries to capture the inferential properties of the natural language words that correspond to the connectives of propositional logic – the logic, for example, gives up on the truth-functionality of entailment: the conditional is understood as *directional*, meaning that, for example, $p \rightarrow q$ does not imply $\neg q \rightarrow \neg p$. An "if...then..." statement thus only allows inferences from information about $p$ to information about $q$ but not in the other direction. By defining logical connectives in a cognitively more plausible way, this logic can better account for human performance on reasoning tasks, e.g. Wason's selection task that was introduced above. Braine's work can quite easily be classified as a work of natural logic – it investigates natural language expressions (that correspond to the connectives of propositional logic) and their inferential properties in a way that fits human performance. The motivation for this can thus be said to be *cognitive*, disregarding computational or even metalogical properties completely.

Endrullis & Moss (2015) discuss a proof system that is concerned with the inferential properties of MOST (interpreted as "strictly more than half") – the fragment that their logic models is given by sentences of the form "All $X$ are $Y$", "Some $X$ are $Y$", and "Most $X$ are $Y$". They proof that their system is sound and complete and provide a proof search algorithm, highlighting the low complexity of their logic. Their approach emphasizes the importance of the motivation that leads to the use of a natural logic: while ours is rather cognitive, they emphasize logical, algorithmic, and complexity results. This case proofs to be an interesting one as its classification as natural logic is not entirely clear – the proof system however operates on natural language sentences and tries to capture the inferential properties of MOST from a computational viewpoint but is trading these additional computational insights for a loss in cognitive plausibility. The proof system does for example not contain any manifestation of existential import – we will later argue against the cognitive plausibility of this. Nevertheless, this logic captures some inferential properties of ALL, SOME and MOST.

Another approach to MOST was offered by Strößner (2017). She introduces a probabilistic entailment rule into a modal predicate logic that can be read as "therefore, probably". A statement of the kind $\text{MOST}(A, B)$ then allows for a probabilistic inference that an individual of kind $A$ is also of kind $B$. This assigns probability not to the resulting proposition, but to the inference. With this new logical constant, given the statement "most semanticists are linguists", one can make the probabilistic inference that a given semanticist is "therefore, probably" a linguist. This logic specifies the inferential properties of the determiner MOST in a probabilistic way but gives up on natural language representation – we can thus not call this approach a natural logic as it loses its most characteristic property, its connection to natural language. Furthermore, one can call into question the flexibility of this model: similarly as the probability heuristics model with its probabilistic semantics that was introduced above, this model cannot deal with cardinal quantifiers.

Keenan (2003 and 2004) studies inferential properties of quantifiers that are related through *complement* and *postcomplement* relations. In that vein, the quantifier NO STUDENT is the postcomplement of EVERY STUDENT because we can infer ″every student passed“ from ″no student did not pass“ and vice versa. Keenan thus states that the semantic properties of quantifiers allow for inferences on their duals: if a quantifier $Q_1$ can be characterized as the dual of another (e.g. $Q_1 = \neg(Q_2\neg)$), this allows for some inferences between them. Highlighting the inferential relationships between quantifiers of all kind operating on natural language representations makes for a natural logic – Keenan however offers no additional insights, e.g. alignment with empirical data or testable predictions.

Finally, we wish to remark that the ″natural“ approach is expanding: Seuren (2010) offers an overview over natural set theory approaches that, similarly to Braine's propositional natural logic (above) tries to give a more plausible meaning to set-theoretical operations such as ″$\cup$“ and ″$\cap$“. We will however not further review this approach here and continue our way toward a bigger picture of natural logic. Furthermore, Bowman *et al.* (2015) use neural networks to learn distributional semantics word representations and investigate whether they, by that, learn inferential relationships between words, e.g. that from something being a turtle follows that it is not a chair.

## 3.2 A Natural Logic Roadmap

We have seen different motivations for using natural logic models. Here we will try to carve out the main ones and then propose a natural logic methodology. As we have seen, one of the unifying characteristics of natural logic is that it operates on natural language – as a corollary of this, a natural logic model operates on representations that preserve the essential properties of natural language representations. With this comes great expressibility and thereby flexibility (as opposed to first order logic representations).

We have furthermore seen that natural logics vary in their motivations - while one can focus on cognitive plausibility – as Braine (1978) did –, one can also focus on metalogical properties or even automatic theorem provers – as Endrullis & Moss (2015) did. Accordingly, one might want to have one's natural logic to be informed by the psychology of reasoning and then formulate appropriate reasoning rules. We will give a quick overview over the most important points. We can now formulate, in a concise way, which factors proof to be important for a natural logic.

**Natural Logic Roadmap**

1. identify suitable language fragment and core semantic or syntactic principles that inferences can be build on, e.g. monotonicity
2. formulate goal (cognitive, logical, exploratory)
3. formulate appropriate inference rules – if corresponding to goal, let them be informed by empirical results

4. if corresponding to goals: formulate complexity measure that captures the cognitive difficulty of inferences on the fragment
5. evaluate logic according to initial goals

Our logic is concerned with what is a good inference and how hard it is to get there – we will later talk about the philosophical implications of seeing the inferential characteristics of natural language expressions as their defining ones. We are convinced that natural logics should operate under a simplicity constraint: inferences should be added to a logic only if they are useful – one should aim for the *minimal logic* that models the fragment in question.

# 4 Generalized Quantifiers

We will now shift our attention towards the other topic that dominates the title of this work: generalized quantifiers. These are known as such because, historically speaking, their formal, model-theoretic study was an extension of the study of the existential (∃) and universal (∀) quantifiers. The study of quantification is at the intersection of linguistics, logic, mathematics and psychology and is – from a ″natural“ viewpoint – especially interesting because, as Peters & Westerståhl (2006) note

> ″Quantifiers are one of very few expressive devices of language for which it is known how to break out of the circle of language and explain what a word means other than essentially in terms of other words' meanings.“ (Peters & Westerståhl 2006, Preface)

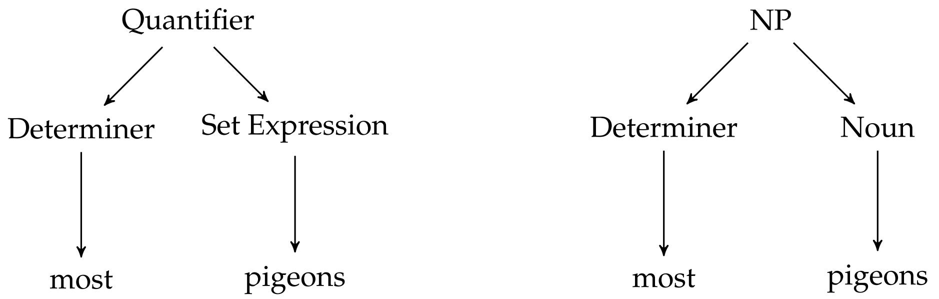
This indicates that interpretation is unproblematic – we will however later see examples that poof such an assumption to be too optimistic. But this and their relatively straight-forward model-theoretic semantics (together with further background in Appendix A) makes them attractive for the formal study of natural language. We are however not mainly interested in their extensional but inferential properties, as those invite the use of natural logics.[10]

The introduction of modern studies of generalized quantifiers to linguistics is usually attributed to Barwise & Cooper (1981) and is mainly motivated by the inadequacy of first order logic to study aspects of natural language quantification – both existential and universal quantification are rather atypical from a natural language perspective. On

[10] Let us first fix a technicality: there is a large variety of quantifiers that are usually classified in types <$n_1, ..., n_k$> with $k \in \mathbb{N}\backslash 0$ and $n_i \in \mathbb{N}\backslash 0$. We will however restrict ourselves to type <1,1> quantifiers and also refer to them as *binary* quantifiers. The appendix A offers some technical background on this and how the definitions of quantifiers offered in a bit are properly derived from set-theoretic semantics. The restriction to type <1,1> quantifiers essentially means that we will only use quantifier expressions that follow the syntactic form of a determiner combined with a set expression (see below). When talking about quantifiers, we will henceforth mean type <1,1> quantifiers, unless stated otherwise.

top of that observation, there is the well known result that some quantifiers (e.g. MOST) cannot be expressed in terms of first order logic at all (Barwise and Cooper 1981, 160).[11]

A syntactic discussion of binary quantifiers can start by clarifying the relationship between them and determiners in noun phrases (NP) in the English language. While the expressions mostly associated with generalized quantifiers are of the kind MOST, MORE THAN HALF and MANY, those are determiners, not quantifiers. It takes the combination of a determiner with a set expression to obtain a generalized quantifier (Barwise and Cooper 1981), reminiscent of the structure of noun phrases in the English language:



MOST is thus not a quantifier but a determiner – MOST PIGEONS on the other hand is a quantifier. This analysis fits well into our natural logic perspective: if our formal treatment can mirror language, it is considerably more adequate than first order logic representations, which tend to be linguistically quite different from their natural language expressions (while still being logically equivalent, see examples below). Barwise and Cooper (1981, 164 ff.) call this ”the notorious mismatch between the syntax of noun phrases in a natural language like English and their usual representations in traditional predicate logic“, which they illustrate using the following example. Analyzing the following sentences that consist of a noun phrase followed by a verb phrase (left-hand side)

(i) Harry flies fly(h)
(ii) Some pigeon flies $\exists x(pigeon(x) \wedge fly(x))$
(iii) Every pigeon flies $\forall x(pigoeon(x) \rightarrow fly(x))$
(iv) Most baby-pigeons fly (no FOL translation of this sentences)

one sees that the NPs here (Harry, some pigeon, every pigeon, most baby-pigeons) likely belong to the same syntactic category but translate to first order logic quite differently (right-hand side), introducing differences, where there are none in natural language. Thus, disregarding the syntactic similarities of these sentences in natural language, the first-order representation of (ii) and (iii) does not contain a representation of the NP (while (i) does) and (iv) is not even translatable to first order logic. Even

[11] On a sidenote, it is not only first order logic that showed difficulties with formalizing MOST. Similarly, van Benthem's approach of semantic automata, i.e. using finite state machines to capture and formalize the meaning of quantifiers, cannot be used to formalize MOST (van Benthem 1986, 153).

worse, (ii) can at best be translated as ”something is a pigeon and flies“, which might be logically equivalent, but proves linguistically different. First order logic's *notorious mismatch* thus calls for a more adequate form of representation. As a corollary of these considerations, we can conclude that first order logic is insufficient for a study of inferences involving generalized quantifiers. This provides with the motivation to use the natural logic approach: natural language representations are both flexible and powerful enough to be cognitively plausible, whereas first order logic is not. Let us now look at the definitions of some quantifiers – how they fit into a model-theoretic setting is explained in Appendix A.

**Some Generalized Quantifiers**
If $Q(A, B)$ is a quantifier, we can usually define it by only referring to the two sets $A$ and $B$. Here are some quantifiers that we will be using:

ALL$(A, B) \Leftrightarrow A \subseteq B$

SOME$(A, B) \Leftrightarrow A \cap B \neq \emptyset$

NO$(A, B) \Leftrightarrow A \cap B = \emptyset$

NOT ALL$(A, B) \Leftrightarrow A - B \neq \emptyset$

MOST$(A, B) \Leftrightarrow |A \cap B| > |A - B|$

FEWER THAN THREE$(A, B) \Leftrightarrow |A \cap B| < 3 \Leftrightarrow |A \cap B| \leq 2$

MORE THAN TWO$(A, B) \Leftrightarrow |A \cap B| > 2 \Leftrightarrow |A \cap B| \geq 3$

AT LEAST THREE$(A, B) \Leftrightarrow |A \cap B| > 2 \Leftrightarrow |A \cap B| \geq 3$

AT MOST TWO$(A, B) \Leftrightarrow |A \cap B| < 3 \Leftrightarrow |A \cap B| \leq 2$

TWO$(A, B) \Leftrightarrow |A \cap B| = 2$

We will henceforth shorten the cardinal quantifiers as to not explicitly write any numbers, e.g. AT MOST instead of AT MOST THREE. Quickly extending our syntactic discussion above, we note that a multitude of grammatical structures allow for quantification: quantification expressions can themselves be NPs (”EVERYTHING is going down“) or adverbs (for a detailed discussion of this, see Peters & Westerståhl 2006, 4 ff.). The syntactic forms of quantification furthermore vary from language to language. Peters & Westerståhl (2006, 11) for example note that the North American language of Strait Salish is known *not* to use determiners as to express quantification. This fact is of significance for the study of generalized quantifiers – we can however not account for it in this work. As the motivation for our work is a cognitive one, we rely on the availability of empirical data to evaluate our model. To the best of our knowledge, the only suitable data available is in English.

Another justification for the limitation to English determiners is the fact that the road we chose is one less traveled than many others – to compensate for this, we need to limit the scope of our investigation.

## 4.1 Characterizing Quantifiers: Montonicity and Symmetry

With our attention successfully restricted, we can now go on and take a closer look at the formal properties of the quantifier expressions that caught our interest. The formulation of a natural logic for reasoning with generalized quantifiers will hinge on the description of inferences through monotonicity- and symmetry-properties.

### 4.1.1 Monotonicity

Some quantifiers will be hard to characterize in terms of their monotonicity properties – and MOST will prove to be especially trying. Let us look at the definition and see what we can do with it (we will largely follow the presentation and notation of Peters & Westerståhl 2006).

**Monotonicity**
A function $F$ is called monotone *increasing* relative to an ordering $\leq_1$ of the arguments and an ordering $\leq_2$ of the values iff

$$x \leq_1 y \Rightarrow F(x) \leq_2 F(y)$$

In terms of generalized quantifiers, "$\leq_1$" is inclusion and "$\leq_2$" implication, thus stating that $Q$ is *left-side monotone increasing* (or *left-side upward monotone*) iff

$$A \subseteq A' \Rightarrow Q(A, B) \rightarrow Q(A', B)$$

and *left-side monotone decreasing* (or *left-side downward monotone*) iff

$$A' \subseteq A \Rightarrow Q(A, B) \rightarrow Q(A', B)$$

Analogous, $Q$ is called *right-side monotone increasing* (or *right-side upward monotone*) iff

$$B \subseteq B' \Rightarrow Q(A, B) \rightarrow Q(A, B')$$

and *right-side monotone decreasing* (or *right-side downward monotone*) iff

$$B' \subseteq B \Rightarrow Q(A, B) \rightarrow Q(A, B')$$

Monotonicity inferences directly follow from this definition, which in some ways extends beyond binary quantifiers. Quantifiers can be increasing in one of the arguments while being decreasing in the other. Some quantifiers like MOST even are right upward monotone, while being neither left upward nor left downward monotone. We will get to this later and first look at some examples of monotonicity and inferences based on monotonicity.

**Monotonicity Profiles**
As a shorthand for the monotonicity properties of a quantifier $Q$, we will talk about its *monotonicity profile*. As an example, instead of stating that a quantifiers is left-side downward monotone and right-side upward monotone, we will say that its monotonicity profile is ↓↑. To make the monotonicity profile of a quantifier explicit, we will often also write $\downarrow Q \uparrow$. If a quantifier does not have any monotonicity properties on a side, we will write a dot instead. Examples are

↓ ALL ↑

↑ SOME ↑

↓ NO ↓

↑ NOT ALL ↓

·MOST ↑

↓ FEWER THAN THREE ↓

↑ MORE THAN TWO ↑

↑ AT LEAST THREE ↑

↓ AT MOST TWO ↓

·TWO·

### 4.1.2 Symmetry and Conversion

Another property of generalized quantifiers that allows for inferences is symmetry. But while most quantifiers that we have seen so far has some kind of monotonicity properties that allow for inference, most quantifiers do not when it comes to symmetry. Let's look at the definition.

**Symmetry and Conversion**
A type quantifier $Q$ is called *symmetric* if and only if, for all $A$ and $B$,

$$Q(A,B) \Rightarrow Q(B,A)$$

the inference associated with this property is called *conversion*.

The entailment does not need to be bidirectional and already has all generality. This will later allow for the definition of a powerful inference – that will however be restricted to smaller class of quantifiers. Some examples are:

(i) (a) Some pigeons are birds.

(b) Some birds are pigeons.

(ii) (a) No pigeons are philosophers.

(b) No philosophers are pigeons.

Whereas the following examples clearly show the limits of possible inferences on grounds of symmetry, as the pairs of sentences do not imply each other.

(i) (a) All pigeons are birds.

(b) All birds are pigeons.

(ii) (a) Most pigeons are birds.

(b) Most birds are pigeons.

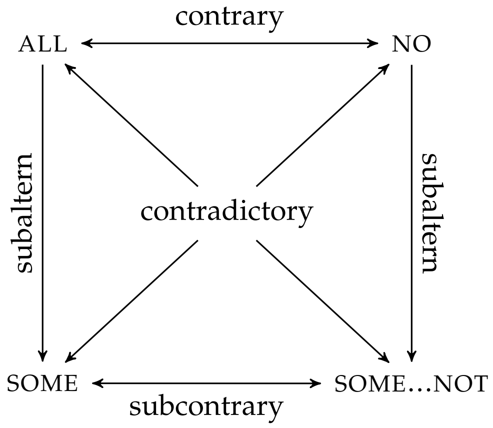
And so on. The truth of the statements only depends on the intersection of two sets, thus allowing for an inference based on symmetry.

## 4.2 Some Remarks on Quantifier Interpretation

Model theorists often make the simplifying assumption that the meaning of basic expressions is fixed by context. We will, in principle, follow this lead here – Barwise & Cooper call this the *fixed context assumption* (Barwise & Cooper 1981, 163) – meaning that we take the interpretation of quantifier expressions to be clear. While this is a strong assumption for context-dependent determiners such as FEW, it is usually taken for granted for most determiners and we will do the same for those in the syllogistic- and the QI-fragments. We will however take a moment to justify their interpretations – especially that of MOST – but notice that the precise interpretation of quantifier expressions is often not as relevant for natural logics as long as their inferential properties are preserved.

### 4.2.1 Aristotelian Quantifiers

Aristotle used the *square of opposition* to organize the syllogistic quantifiers and the relationships that hold between them (Westerståhl 1989, 582).



While the relationships between these quantifier expressions seem relatively clear, we wish to point out the fact that in this picture, ALL implies SOME. This validates the inference

$$\text{exImp} \qquad \frac{\text{ALL}(A,B)}{\text{SOME}(A,B)}$$

that we have already seen above. A simplifying assumption that we make is to equate synonyms: we will consider ALL, EACH, EVERY, etc. to be the same. We will use this point to introduce two further quantifier expressions as semantic primitives. As the Aristotelian SOME...NOT is not usually considered a quantifier, we will define SOME NOT as

$$\text{SOME NOT}(A,B) \Leftrightarrow A - B \neq \emptyset$$

which is defined the same way as NOT ALL introduced above but more similar to the Aristotelian form which is predominantly used in experiments. Analogous, we introduce a primitive for ALL...NOT

$$\text{ALL NOT}(A,B) \Leftrightarrow A \cap B = \emptyset$$

which is the same as NO. We will use these new quantifier expressions extensively when proving syllogisms in Appendix B. The reason why we prefer SOME NOT$(A,B)$ over SOME$(A,$NOT $B)$ is that the latter makes use of a negation with unclear semantics. Note that both ALL NOT, just as ALL, has existential import. Many of the proofs in appendix B will heavily rely on deriving O-propositions from ALL NOT statements, i.e. on ALL NOT's existential import.

### 4.2.2 The Curious Case of MOST

We have seen further above – exemplified with a quote by Peters & Westerståhl – that quantifiers often have straightforward interpretations. We will now consider MOST, which is quite telling concerning the connection between interpretation and inferential properties.

Barwise & Cooper (1981, 163) note that the meaning of MOST depends on the model – whereas the meanings of the logical quantifiers $\forall$ and $\exists$ do not. While their fixed context assumption mitigates this issue by stating that the semantics of the non-logical determiners is contextually fixed, we can also bring forward empirical evidence that supports the interpretation given above that MOST$(A,B)$ is true if and only if more $A$s are $B$s that $A$s are not $B$s.

Pietroski *et al.* (2009) design experiments to inquire the quantifier representations that reasoners form. Using such methodology, one can give empirical bite to semantic distinctions – an endeavor that is highly welcome in natural logics. We have already introduced MOST as a comparative quantifier and their research indicated that this is indeed the cognitively most plausible representation – competing with an understanding relying on bijections, making any counting capacities unnecessary. This corresponds to our earlier assertion that generalized quantifiers should represent sets: the relation ">" holds between the two respective cardinalities of two sets, while the one-to-one correspondence-relation holds between the individuals in those sets (Pietroski *et al.* 2009, 581). This resolves which of the two representations is cognitively more plausible.

Taking a step back, there is a variety of opinions on not only the representation, but also the meaning of MOST. Other researchers found that MOST is actually the superlative of MANY (see for example Kotek *et al.* 2015 and Hackl 2009). In this interpretation, MOST would not be seen as a semantic primitive but a function of the meaning of its parts *many* and *-est*. Hackl (2009) provides two main arguments for this position: firstly, one needs to distinguish comparative and superlative morphosyntax. Proportional quantification that is based on comparative morphosyntax usually has direct opposites (witness MORE THAN HALF versus LESS THAN HALF) while proportional quantification that is based on superlative morphosyntax does not – with MOST being the primary example, as there is no such quantifier as FEWEST (Hackl 2009, 64). According to them, then, MOST is in an entirely different category of proprotional quantifiers than MORE THAN HALF. Secondly, they provide empirical evidence that subjects in experimental settings choose different verification strategies for MOST and MORE THAN HALF while maintaining the same model-theoretic semantics. From the perspective of natural logics, we can observe that this distinction does not matter: MANY and MORE THAN HALF have the same right-side monotonicity properties (both are right-side upward entailing). Discussing the interpretation of MOST is thus not prevalent from an inferential viewpoint – yet. We will later see that MORE THAN HALF allows for more inferences beyond monotonicity (Similarly, the logic proposed by Endrullis & Moss (2015) that was introduced shortly above pivots around the interpretation of MOST as MORE THAN HALF). We will thus stick to the somewhat orthodox interpretation of MOST as MORE THAN HALF.

### 4.2.3 Cardinal Quantifiers

The last kind of quantifier that we wish to look at in this chapter is cardinal quantifiers. While we think that the interpretations of cardinal quantifiers are relatively straightforward, there are still some remarks to be made. Geurts *et al.* (2010) highlight the difference between superlative and comparative quantifiers, e.g. between the two sentences

(i) Berta had at least three beers.

(ii) Berta had more than two beers.

Note that we introduced the interpretations of these two kinds of sentences to be the same above. Geurts *et al.* (2010) however argue that superlative and comparative quantifiers give raise to different inferences, highlighting that superlative quantifier expressions (AT LEAST, AT MOST) are cognitively more difficult than comparative ones

(MORE THAN, FEWER THAN). So, while their model-theoretic interpretation remains the same, of the two inferences

(i) $\dfrac{\text{Berta had three beers.}}{\text{Berta had more than two beers.}}$ (ii) $\dfrac{\text{Berta had three beers.}}{\text{Berta had at least three beers.}}$

only the first one is a *good* inference as (ii) opens up the possibility of Berta having more than three beers – which the premise clearly states she did not have. Geurts *et al.* thus reject the two quantifiers having the same meaning for pragmatic reasons. Our resolution to this conundrum will again be to emphasize that it is not important from our viewpoint: important is that both interpretations do not differ in their monotonicity properties – which they do not. We will generally uphold a simplicity constraint stating that interpretation-dependent rules should only be introduced when they are *useful* (e.g. exImp) and not just when they are *possible*.

## 4.3 Quantifier Iteration

Our technical treatment of iterated generalized quantifiers will be rather quick.[12] Consider a sentence like ”MOST A played against AT LEAST THREE B“ which can be formalized as follows:

$$Q_1, Q_2\phi(A, B)$$

Is it clear that we cannot make any symmetry-based inferences as the symmetry of the statement depends on the relation $\phi(\cdot)$ just as much as on the quantifier expressions used. Crucially, however, monotonicity based inferences are still possible. We will introduce the notion of a combinatorial monotonicity profile, an extension of the notion of monotonicity profiles introduced above, to fully appreciate this.

**Combinatorial Monotonicity Profiles (CMP)** Similar as single quantifiers, most iterated quantifiers put their arguments in positions that allow for monotonicity inferences. Iterated quantifiers then have a monotonicity profile similar to those of single quantifiers. So, for example,

$$\downarrow Q_1, Q_2 \uparrow$$

means that the iteration of $Q_1$ and $Q_2$ puts the first argument in a downward entailing position and the second argument in an upward entailing position.

The work of Geurts and van der Silk (2005, 108) suggests that the combined (right-side) monotonicity profile of two iterated quantifiers depends only on their respective right-side monotonicity properties, but this is not the case. The interaction that they

[12] Strictly speaking, iterated quantifiers are a type of polyadic quantifiers – so if $Q_1$ and $Q_2$ are both type <1>, their iteration is type <2>. Technically, iterating two quantifiers gives raise to one single quantifiers that can have semantical properties as introduced above. We will however not get too technical here and refer to Peters & Westerståhl 2006, 346ff.

propose is that if $Q_1$ is right-side downward entailing, this reverses the direction of entailment of the second quantifier. This suggests the following interaction that is reminiscent of how subtraction and addition interact in arithmetics:

|   | ↑ | ↓ | · |
|---|---|---|---|
| ↑ | ↑ | ↓ | · |
| ↓ | ↓ | ↑ | · |
| · | · | · | · |

Where a point indicates that a quantifier expression has no monotonicity properties on the relevant side (e.g. the left side of MOST). Consider however the following examples:

(i) ↓NO↓ A played against ↑SOME↑ B. CMP ↓ NO, SOME ↓
(ii) ↓NO↓ A played against ↓ALL↑ B. CMP ↓ NO, ALL ↑
(iii) ↓NO↓ A played against ·MOST↑ B. CMP ↓ NO, MOST·
(iv) ↓ALL↑ A played against ↓ALL↑ B. CMP ↓ ALL, ALL ↓

While (i) conforms to this matrix and puts B into a downward entailing position, (ii) does not – the second argument is still in an upward entailing position. Even worse so, (iii) does not lead to any inference, as the second argument is neither in an upward or downward entailing position (the inference would require more information about the sets involved). As another counterexample, (iv) involved two right-side upward monotone quantifiers that together put their second argument in a downward-entailing position.

The iteration scheme that Geurts and van der Silk (2005) suggest for all quantifiers is thus only working for the combinations of quantifiers in their fragment. While this does not interfere with our plan of creating a complexity measure for inference rules, it *does* interfere with our plan of creating general inference rules. With the monotonicity switch indicated in the scheme above not being valid for all combinations of quantifiers, we cannot create general rules that account for all quantifier expressions. This does not mean that no monotonicity-inferences are possible but that they are not a straightforward generalization of the case of single quantifiers. We can however state that left-side monotonicity properties are unproblematic: those are directly inherited from the first quantifier.

The semantics of quantifier-expressions will not only give raise to the inference rules, but will also partially be responsible for their weight-assignments: as we can see in the algebra above, some quantifier iterations give raise to a change in the directionality of monotonicity-entailments. We can already speculate that this requires additional processing.

# 5 A Natural Logic for Reasoning with Generalized Quantifiers

It is now time to present our natural logic. We will creatively call the entirety of inference rules and weight assignments natural quantifier logic (NQL) and wish to remind that this logic can only make predictions about *good* inferences and offers no hypotheses on when and why reasoners make bad inferences (although we will see later on that such a hypothesis is possible in some cases). First, we will fix vocabulary and syntax for our model.

Vocabulary

- basic terms: A, B, C ... (large letters)
- Binary quantifier expressions: ALL, SOME, NO, SOME NOT, MOST, FEWER THAN, MORE THAN, AT LEAST, AT MOST
- Arrows indicating monotonicity properties and brackets: ↑, ↓, ↑↑, ↑↓, ↓↑, ↓↓ ), (

Syntax

- If $A$ and $B$ are basic terms and $Q$ is a quantifier, then $Q(A,B)$ is a sentence
- If $A$ and $B$ are basic terms and $\phi$ is a relation, then $\phi(A,B)$ is a formula
- If $\phi(A,B)$ is a formula and $Q_1$ and $Q_2$ are binary quantifiers, then $Q_1, Q_2\phi(A,B)$ is a sentence

We further need inference rules, which we will now present step by step.

## 5.1 Inference Rules

We will start with reasoning on single quantifiers, make our way toward quantifier iteration an then extensively motivate our weight assignments that assign each inference rule a complexity which, as we will later show, correlates with cognitive difficulty..

### 5.1.1 Single Quantifiers

**Inference Rules for Single Quantifiers**
The following inference rules allow for proving all syllogisms that are valid in predicate calculus and / or Aristotelian logic (and more).

$$\text{Mon}\uparrow \qquad \frac{\begin{array}{c}\mathrm{Q}\uparrow(A,B)\\ \textsc{all}(B,C)\end{array}}{\mathrm{Q}\uparrow(A,C)} \qquad\qquad \text{Mon}\downarrow \qquad \frac{\begin{array}{c}\mathrm{Q}\downarrow(A,B)\\ \textsc{all}(C,B)\end{array}}{\mathrm{Q}\downarrow(A,C)}$$

$$\uparrow\text{Mon}\qquad \frac{\begin{array}{l}\uparrow \mathrm{Q}(A,B)\\ \textsc{all}(A,C)\end{array}}{\uparrow \mathrm{Q}(C,B)} \qquad\qquad \downarrow\text{Mon}\qquad \frac{\begin{array}{l}\downarrow \mathrm{Q}(A,B)\\ \textsc{all}(C,A)\end{array}}{\downarrow \mathrm{Q}(C,B)}$$

$$\text{Conv}\qquad \frac{\mathrm{Q}_s(A,B)}{\mathrm{Q}_s(B,C)} \qquad\qquad \text{pConv}\qquad \frac{\textsc{no}(A,B)}{\textsc{all not}(A,B)}$$

$$\text{exImp}\qquad \frac{\textsc{all}(A,B)}{\textsc{some}(A,B)}$$

With $Q_s$ denoting any symmetric quantifier (NO, SOME, all cardinal quantifiers, etc.) and $Q\uparrow$, $Q\downarrow$, $\uparrow Q$, and $\downarrow Q$ are binary quantifiers with the indicated monotonicity properties..

Note that with its use of left-side monotonicity properties and the generality in quantifier assignment, this already extends far beyond Geurts' model. While this introduces all monotonicity-based inferences, we will later-on make one more distinction between them: those that do and those that do not feature a negative context – with the latter proving to be harder. Monotonicity inferences involving a SOME NOT or ALL NOT are considerably harder. We wish to furthermore emphasize that there is no conversion rule for SOME NOT.

### 5.1.2 Inference Rules for Iterated Quantifiers

**(Right-Side) Monotonicity Based Inference Rules**
The following inference rules account for all of the QI-fragment with its limitation on the second quantifier.

$$\text{Mon}\uparrow\uparrow\qquad \frac{\begin{array}{l}Q\uparrow Q_M\uparrow \phi(A,B)\\ \textsc{all}(B,C)\end{array}}{Q\uparrow Q_M\uparrow \phi(A,C)} \qquad\qquad \text{Mon}\uparrow\downarrow\qquad \frac{\begin{array}{l}Q\uparrow Q_F\downarrow \phi(A,B)\\ \textsc{all}(C,B)\end{array}}{Q\uparrow Q_F\downarrow \phi(A,C)}$$

$$\text{Mon}\downarrow\uparrow\qquad \frac{\begin{array}{l}Q\downarrow Q_M\uparrow \phi(A,B)\\ \textsc{all}(C,B)\end{array}}{Q\downarrow Q_M\uparrow \phi(A,C)} \qquad\qquad \text{Mon}\downarrow\downarrow\qquad \frac{\begin{array}{l}Q\downarrow Q_F\downarrow \phi(A,B)\\ \textsc{all}(B,C)\end{array}}{Q\downarrow Q_F\downarrow \phi(A,C)}$$

Where $Q$ is any binary quantifiers with the indicated monotonicity properties, and $Q_M$ and $Q_F$ are MORE THAN and FEWER THAN, respectively. These inference schemes hold for all generalized quantifiers with the right monotonicity properties.

These inference schemes make only use of monotonicity properties exactly because the Geurts-fragment is exclusively concerned with monotonicity inferences. We wish to highlight that these inferences are restricted to the QI-fragment. The examples that were introduced at the end of the last chapter show that this approach runs into problems outside of it as the there are special combinatorial properties with our restrictions on $Q_M = \textsc{more than}$ and $Q_F = \textsc{fewer than}$. We will for now concede that, as our goal was to model this specific fragment, this is not further problematic and come back to this issue later. We mentioned above that we cannot introduce any inference analogous to Conv as symmetry for iterated quantifiers depends on the relation $\phi$. We can further extend this to account for left-side monotonicity inferences, whose directionality only depends on the first quantifier expression:

**(Left-Side) Monotonicity Based Inference Rules**
The following inference rules account for all of the QI-fragment.

$$\uparrow\uparrow\text{Mon} \quad \dfrac{\begin{array}{l} Q_1 \uparrow Q_2 \uparrow \phi(A,B) \\ \textsc{all}(A,C) \end{array}}{Q_1 \uparrow Q_2 \uparrow \phi(C,B)} \qquad \uparrow\downarrow\text{Mon} \quad \dfrac{\begin{array}{l} Q_1 \uparrow Q_2 \downarrow \phi(A,B) \\ \textsc{all}(A,C) \end{array}}{Q_1 \uparrow Q_2 \downarrow \phi(C,B)}$$

$$\downarrow\uparrow\text{Mon} \quad \dfrac{\begin{array}{l} Q_1 \downarrow Q_2 \uparrow \phi(A,B) \\ \textsc{all}(C,A) \end{array}}{Q_1 \downarrow Q_2 \uparrow \phi(C,B)} \qquad \downarrow\downarrow\text{Mon} \quad \dfrac{\begin{array}{l} Q_1 \downarrow Q_2 \downarrow \phi(A,B) \\ \textsc{all}(C,A) \end{array}}{Q_1 \downarrow Q_2 \downarrow \phi(C,B)}$$

Where $Q_1$ and $Q_2$ are again being binary quantifiers with the indicated monotonicity properties. These inference schemes hold for all generalized quantifier with the right monotonicity properties.

### 5.1.3 Beyond

We can imagine a variety of ways to take our logic beyond these fragments of natural language. Firstly, we could iterate more quantifiers to account for sentences such as ”MOST philosophers played AT LEAST THREE games of chess against MORE THAN TWO linguists“ – this would however rapidly increase complexity and go beyond our technical treatment of binary quantifiers.

Secondly, while we stated above that MOST does not have any left-side monotonicity properties, this is actually not true – but to fully capture the monotonicity behavior of MOST's first position, we have to make a quick digression and talk about *smoothness*. We will present the result here and let the rest take place in Appendix A.

**Smoothness** A quantifier $Q$ is *smooth* iff the following two conditions hold:

(i) $Q(A,B) \wedge A \subseteq A' \wedge A - B = A' - B \Rightarrow Q(A',B)$

(ii) $Q(A,B) \wedge A' \subseteq A \wedge A \cap B = A' \cap B \Rightarrow Q(A',B)$

This allows us to get a better grasp of proportional quantifiers – they are all smooth (Peters & Westerståhl 2006, 187). The inferences that MOST should allow in a natural logic based on left-side monotonicity properties are thus exactly the ones that in the definition of smoothness. While this completes our treatment of monotonicity, we note that we will not make use of smoothness in our natural logic – making such inferences requires very specific information about the sets $A$, $A'$, and $B$ and not only semantic information.

Thirdly, we could introduce additional rules that capture semantic properties of specific quantifiers. So, for example, MOST in its interpretation as MORE THAN HALF would allow for the inference

$$\frac{\begin{array}{c}\textsc{most}(A,B)\\ \textsc{most}(A,C)\end{array}}{\textsc{some}(B,C)}$$

This extension of the model could however not rely on already existing data and – to the best of our knowledge – could also not rely on psychological literature that informs us about the difficulty of this kind of inference. This is furthermore an interpretation-dependent inference, which we wish to avoid as far as possible

Lastly, we could also pay more attention to the relationships between single quantifiers: akin to the rule that we call exImp, we could also say that MOST implies SOME in a similar way:

$$\frac{\textsc{most}(A,B)}{\textsc{some}(A,B)}$$

We could define this kind of inference for a variety of quantifiers but would relatively quickly run into difficulties regarding their respective interpretations: recalling our section on cardinal quantifiers above, we know that inferences that rely on the interpretation of a quantifier expression soon run into possible counterexamples. The model as it stands right now relies mostly on the monotonicity properties of quantifiers and not their interpretation – extending it in the last two ways mentioned here would make it dependent on the adequacy of the interpretations that we offer. And there is by no means a unique viewpoint on this in the literature. We will thus leave the model as it is and focus on weight-assignments and predictions. We furthermore think that a simplicity constraint is appropriate demanding that we only add quantifier-specific inference rules if it is *useful* (as exImp is necessary to account for all valid syllogisms), not just when it is *possible*.

### 5.2 Complexity: a Weighted Number of Reasoning Steps

We will now motivate a complexity-measure for this logic. There are three crucial ideas:

- the number of reasoning steps from premises to a conclusion is the length of its proof in natural logic
- some reasoning steps are harder than others
- we can account for this difference in difficulty by assigning a weight (cost) to inference rules, summing up to different costs for different proofs

We will try to motivate the assignment of weights to inference rules with either semantical or psychological considerations and see that a major difficulty is to find a non-arbitrary relationship between the difficulties of inferences that do not have any telling semantic relationship, e.g. monotonicity- and symmetry-grounded inferences or monotonicity-inferences of single and iterated quantifiers. In those cases, we will also rely on a good fit to the data. We can immediately note two things: first, monotonicity inferences are easy (see for example syllogisms like AA1A and ↑↑-inferences in the QI-fragment that only involve monotonicity). Secondly, monotonicity inferences are hard (see for example ↓↓-inferences in the QI-fragment that only involve monotonicity). We will thus spend much time talking about the variations of monotonicity inferences and try to zoom in on their differences and commonalities. One of the main results of our exercise in weight-assignment will be that while existential import should be part of our logic, inferences based on existential import are quite unlikely to be made, one reason being that the involvement of the quantifier expression ALL means that they often have to compete for reasoners attention with the somewhat more exciting monotonicity-based inferences.

The number of reasoning steps is operationalized as the length of proof from a set of premises to a conclusion. The assignment of weights is especially important since all inferences in the QI-fragment have length 1 while showing large variation in cognitive difficulty. Before we introduce our complexity-measure, we will talk about ideas on the difficulty of reasoning with quantifier expressions that will *not* factor into our model.

As in our discussion of Geurts' (2003) logic, we noted that they weight proofs in the syllogistic fragment that involve an O-proposition (SOME...NOT) as cognitively more difficult than others. This approach gives raise to a better fit to the data, but we will not consider this here: our complexity measure should be a weighted number of reasoning steps that hence only weights *inferences*, but not *propositions*. Weighting propositions is a step away from our focus on proof-length. While Geurts argues that such a proposition is harder because it contains a negative, Newstead (2003, 195) points out that with the same argument, proofs that involve an E-proposition (NO) should be harder as well. One could however argue that this is not the same kind of negation and that NO is a primitive while SOME...NOT is not. From such a viewpoint, it would truly be the alteration of the primitive SOME that would cause additional cost. The experimental evidence regarding this issue in fact points into no clear direction, with Hardman & Payne (1995) and Roberts *et al.* (2001) claiming that people are not in any way more reluctant to draw O-conclusions than other types. We will later propose a way that accounts for Geurts' point but at the same time maintains focus on inferences and proof length, this stance however demands fr more testing.

Similarily, Szymanik & Zajenkowski (2010) conducted an experiment to confirm their prediction that the use of computational resources has an impact on performance of cognitive tasks. Based on the theory of semantic automata (van Benthem 1986), they predicted that quantifiers that are recognized by acyclic (i.e. without any loops) finite automata (first order quantifiers), quantifiers that are recognized by finite automata (e.g. quantifiers that express parity, such as AN EVEN NUMBER OF), and quantifiers that can only be recognized by push-down automata (proportional quantifiers such as MOST) make cognitive tasks harder: finite automata are without memory, while push-down automata have a limited memory – proportional quantifiers thus require more cognitive ressources. Their predictions were confirmed by an experiment in Polish. Accordingly, we could weight inferences on proportional quantifiers more than others.

A similar distinction between quantifiers was done by Szymanik & Thorne (2017), who conducted a frequency analysis of generalized quantifiers on a corpus derived from Wikipedia and found a correlation between frequency and complexity. One first immediate insight of their frequency analysis is that AT MOST appeared only 619 times, while NO with its 464′755 appearances one of the most frequent quantifiers in these corpora. While it is certainly a plausible proposition that the processing of more frequent quantifiers is easier, we will also dismiss this here, the reason being – similarly to the issue of O-propositions – that we wish to create a measure of difficulty of *inferences*, not *expressions* or *propositions*. Though it might yield an explanation why inferences involving NO proof to be so much easier than inferences involving AT MOST in the QI-fragment. Similarly, we will ignore the interactions between quantifiers except for those that are grounded in monotonicity. Let us look at those first.

### 5.2.1 Monotonicity I and II: Directionality and Harmony

Recall that the four combinations of right-side monotonicity properties for two iterated quantifiers are ↑↑, ↑↓, ↓↑ and ↓↓. This immediately allows for the clarification of two points: how many of the quantifiers involved ″go up‟ and whether both have the same directionality. We propose a cost-based system in which less favorable inferential properties add cost: *directionality* raises cost according to how many downward monotone quantifiers are involved (normalized such that all values are between 0 and 1 – the increased difficulty of downward inferences was argued by Geurts & van der Silk (2005, 104) and Clark (1973 and 1974) – and *harmony* gives raise to cost when the two quantifiers do not have the same directionality.

Note that while harmony is a yes-or-no question, directionality is not – two sequential quantifiers either exhibit harmony or not, but downward-directionality comes

**Table 6:** Combinatorial monotonicity profiles (CMP) and the quantifier pairs that correspond to them in the QI-fragment

| CMP | Quantifier Pairs |
|---|---|
| ↑↑ | (every, more than), (most, more than),<br>(at least, more than), (some, more than) |
| ↑↓ | (every, fewer than), (most, fewer than),<br>(at least, fewer than), (some, fewer than) |
| ↓↑ | (at most, more than), (no, more than) |
| ↓↓ | (at most, fewer than), (no, fewer than) |

in different magnitudes. Furthermore, non-iterated quantifiers have no costs related to harmony, as there cannot be any dissonance in their upward- and downwardness. We can now make a more perspicuous overview over what pairs of quantifiers in the quantifier-iteration fragment exhibit harmony and directionality (table 6).

Recalling table 4, we can see that the significant difference connected to the use of AT MOST cannot be explained by the semantic criteria used here: inferences involving AT MOST are – from a perspective of monotonicity – not different from the ones that involve NO in the major premise. This motivates our following investigation into the informativeness of generalized quantifiers.

### 5.2.2 Monotonicity III: a Hierarchy of Informativeness

Our conception of informativeness is a semantic one. Informativeness is another recurring topic in the discussion of generalized quantifiers – and there are different approaches to its conceptualization. Oaksford *et al.* (2002), for example, use experimental methods to fix a hierarchy of informativeness of generalized quantifiers: ALL > MOST > SOME > FEW > NONE > SOME...NOT, as e.g. a statement involving ALL is less probable than the same statement with SOME. We think, however, that this approach is misleading.

From a semantic point of view, ALL and NO are the most informative. There are various different ways to say why this is the case: Katsos *et al.* call it the *totality* of ALL and NO, while our approach states that ALL and NO are somewhat on top of a monotonicity-based semantic food chain, that is, more informative quantifiers allow for inferences that less informative ones do not.

**Semantic (Right-Side) Informativeness**
The quantifiers ALL and NO are *informative*, i.e. they allow for the inferences

$$\text{ALL}(A, B) \Rightarrow Q \uparrow (A, B)$$

$$\text{NO}(A, B) \Rightarrow Q \downarrow (A, B)$$

for any right-side upwards monotone quantifier $Q \uparrow$ and for any right-side downwards monotone quantifier $Q \downarrow$, respectively.

Case in point is the existential-import rule introduced further above. This discussions also mirrors the classification of Aristotelian quantifiers into universal and particular ones: the universal ones prove to be more informative and imply their subaltern particulars. We state that inferences using informative quantifiers are less costly. Note that the notion of semantic (right-side) informativeness is sufficient for the QI-fragment, as it is only concerned with right-side monotonicity inferences.

### 5.2.3 Monotonicity IV: Negativity and the Monotonicity-Rollercoaster

A further factor that concerns the the case of iterated quantification is that monotonicity may change its direction. In a sentence with two iterated quantifiers of the form $Q_1 Q_2 \phi(A, B)$, formed according to the rules of the QI-fragment, which might state "All $A$s played against at least two $B$s", the second quantifier $Q_2$ puts the second argument $B$ in either an upward or downward entailing position. However, if the first quantifier $Q_1$ is downward entailing, it turns the entailment direction of the second quantifier upside down. This monotonicity-rollercoaster requires additional processing – thus, if one has two right-side monotone quantifiers $Q_1 \uparrow$ and $Q_2 \downarrow$, the two sentences

(i) $Q_1 \uparrow Q_2 \downarrow \phi(A, B)$

(ii) $Q_2 \downarrow Q_1 \uparrow \phi(A, B)$

both have their second argument in a downward entailing position, but (ii) is harder than (i) because it requires the additional processing of changing the direction of entailment. A monotonicity inference of the kind in (ii) is thus harder than an inference of the kind in (i). In the same way, in

(iii) $Q_1 \uparrow Q_1 \uparrow \phi(A, B)$

(iv) $Q_2 \downarrow Q_2 \downarrow \phi(A, B)$

(iii) requires no change of direction while (iv) does. This change of direction adds additional cost. We can observe a similar phenomenon occurring independent of iteration. Consider syllogism AO2O:

$$\text{AO2O} \qquad \frac{\begin{array}{l}\text{ALL}(C, B)\\ \text{SOME NOT}(A, B)\end{array}}{\text{SOME NOT}(A, C)}$$

This inference only involves one application of Mon↓. But SOME is right-side upward monotone – the NOT before the second argument however changes its directionality. This adds additional cost and justifies a difficulty-based distinction between monotonicity inferences.[13]

### 5.2.4 Conversion and Pseudoconversion

As mentioned above, some generalized quantifiers are symmetric, i.e. they allow for the inference

$$Q(A, B) \Rightarrow Q(B, A)$$

to be made. Key examples are SOME and NO, but also cardinal quantifiers of different sorts: AT LEAST X, FEWER THAN Y, EXACTLY THREE, etc. As noted above, Geurts (2003) claims that while pConv has small cognitive cost, Conv itself has none. This was criticized by Newstead (2003) who is especially reluctant to accept the low cognitive cost of pConv. In absence of any empirical evidence on this (Newstead 2003, 195), we will settle somewhere on the middle ground: Conv is *not* without any, but with very small cognitive cost, so is pConv. We concede that this procedure is somewhat ad hoc but blame this on the lack of specific experimental data. Our stance will later however be partially justified by an increased fit to the data on syllogisms.

### 5.2.5 Existential Import

Peters & Westerståhl (2006, 124) note that native speakers of English usually take statements including ALL or EVERY to imply that they do have instances, i.e. that the restriction $A$ in $\text{ALL}(A, B)$ is actually non-empty. Similarly, Geurts (2007, 258) notes that this assumption, call it the *existential import of universal statements*, has been unchallenged for over 2000 years – and that it has been part of all psychological investigations into syllogistic reasoning. While this might just be due to the tradition manifested in Aristotle's square of opposition, there are also some proper empirical arguments to make this case. On the empirical side, we can note the experiment of Rips (1994), in which 65% of the participants endorsed the argument

$$\frac{\begin{array}{l}\text{EVERY } A \text{ is } B\\ \text{EVERY } B \text{ is } C\end{array}}{\text{SOME } A \text{ is } C}$$

which is valid only if EVERY (in our case: ALL) is taken to have existential import (Geurts 2007, 258). Another issue factoring into our decision to include existential import is experimental design: as Khemani & Johnson-Laird (2012, 4) note, subjects are usually informed that individuals of all relevant sorts exist. This means that subject probably do not actively avoid exImp inferences because they do not know whether the variables are empty. It is thus safe to make an exImp inference, and subjects know that.

[13] We make use of the fact that negation is harder to process, e.g. shown in Wason (1961).

In the discussion of such an inference rule, we have to distinguish two different questions: firstly, whether it is cognitively plausible, i.e. whether our logic should feature such a rule. And secondly, how difficult its application is, i.e. how much complexity its application should import into a sequence of reasoning. As for the first aspect, we have already established that our logic should feature the rule above – both from a cognitive and a formal perspective. In light of this, the resolution of the second question might seem somewhat paradoxical.

The work of Katsos *et al.* (2016), while concerned with quantifier acquisition, offers some important insight. As part of their study, they investigate how adults deal with underinformative quantifiers (note that SOME, if ALL proves true, is underinformative by the standards of all theories that were introduced above). For the relevant part of their study, 536 adults (across 31 languages representing 11 language types) were confronted with the following situation: five boxes and five items were presented, between zero and five of the items were situated in boxes. They then had to decide whether a statement using a quantifier was true or false, e.g. that SOME items were in boxes.

In 84% of all cases were the statement was true but underinformative (e.g. using SOME when all items were situated in boxes), the statement was rejected by the participants (Katsos *et al.* 2016, 9246). Furthermore, Chater & Oaksford's meta-analysis (1999) suggests that it is exactly the syllogisms involving the exImp rule that prove hard for people.

This is somewhat paradoxical because other experimental evidence suggests that the existential import of universal statements is cognitively plausible. It seems however that making the actual inference exImp can prove implausible to cognitive agents as it exchanges an informative statement for an underinformative one. We conclude that above inference accentuating existential import should be part of our logic – but that its application is of high cognitive difficulty. This mirrors our stance that while people would often *accept* inferences involving existential import, it is not straightforward to explicitly *make* them oneself, as they substitute an informative with an uninformative statement.[14]

### 5.3 The Difficulty of Reasoning with Single Quantifiers

The inference rules and their weight-assignment will connect all the semantic and psychological dots that we encountered along the way. We will start with single quantifiers

[14] On a pragmatic sidenote, this approach is supported by Grice's (1967) view that a statement such as ALL$(A, B)$ has primarily the task of communicating information, not only the recitation of true facts. As a consequence of this, pragmatically, statements are not only evaluated for their truth – but also, and primarily, for their informativeness and relevance. Confronting a subject in an experiment with the statement ALL$(A, B)$ thus implies that this is informative and relevant. Exchanging this informative statement with a less informative SOME$(A, B)$ is thus against Gricean pragmatic principles (see also Newstead 2003). While a participant might *endorse* the inference in Rips' data (see above), she might not actually make the inference herself.

and continue to relate their monotonicity inferences to their more volatile counterparts on iterated quantifiers. The weights for inferences on single quantifiers are as in table 7. The proofs in Appendix B show that syllogisms that only require one monotonicity-inference are about equally hard – no matter whether the inference is on the left or right side or goes up or down. We will assign a cost of 10 to these. We have already argued that monotonicity inferences involving SOME NOT or ALL NOT are harder than those who do not. The appropriate relationship between the two is indicated by syllogisms OA3O and AO2O who only involve a single application of a monotonicity rule with SOME NOT or ALL NOT – giving the weight $\text{Mon}_N = 30$ to those. We furthermore stated that exImp should be part of the logic but raise very high cost as reasoners are reluctant to draw this inference. While this was supported by related evidence, we will later suggest to further test this. Our stance on Conv and pConv is that they both should involve relatively little cost but not none (as Conv did in Geurts' logic).

**Table 7:** Weights for the inference rules used on the syllogistic fragment. Mon stands for all monotonicity inferences and $\text{Mon}_N$ for monotonicity inferences involving SOMENOT or ALLNOT.

| Inference Rule | exImp | $\text{Mon}_N$ | Mon | pConv | Conv |
|---|---|---|---|---|---|
| Weight | 60 | 30 | 10 | 5 | 5 |

### 5.4 Relating Single and Iterated Quantifiers

Participants in Geurts and van der Silk's (2005) experiments on reasoning with generalized quantifiers rightly judged valid monotonicity inferences to be correct with a mean success varying between 36% and 96%. We will now use our collected remarks on monotonicity above to create weights that do this variation justice.

We noticed that the directionality of the quantifiers involved has an impact on the cognitive difficulty of iterated quantifiers in the sense that downwardness increases cost. Further, cost is increased if the two quantifiers do not have the same directionality (harmony) and if the first quantifier switches the directionality of the second (in the QI-fragment, this is the case whenever the first quantifier is right-side downward entailing). Cost decreases however if one of the informative quantifier expressions is in the major premise (ALL and NOT). All factors are normalized s.t. assigned values are between 0 and 1 (directionality and hierarchy allow for values 0, 0.5, and 1 while hierarchy and switch only allow for values 0 and 1). The results of this are summed up in table 8. Where there are two values, the first one holds when the first quantifier is NO or ALL and the second one if not (note that the second quantifier in the QI-fragment is fixed to MORE THAN or FEWER THAN).

We realized further above that monotonicity inferences on iterated quantifiers need not be harder than those on single quantifiers. We will define the base cost of a mono-

**Table 8:** Costs assigned to combinatorial monotonicity profiles according to upwardness, harmony and switch. "Negative points are gathered that state to which factor an inference relates to the basic cost of 15.

| CMP | Up/Down | Harmony | Switch | Hierarchy | Overall |
|---|---|---|---|---|---|
| ↑↑ | 0 | 0 | 0 | 0/1 | 0/1 |
| ↑↓ | 0.5 | 1 | 0 | 0/1 | 1.5/2.5 |
| ↓↑ | 0.5 | 1 | 1 | 0/1 | 2.5/3.5 |
| ↓↓ | 1 | 0 | 1 | 0/1 | 2/3 |

tonicity inference on iterated quantifiers to be 15, reflecting that they need not be harder but that their difficulty increases faster when the aggravating factors discussed above come into play. The cost of an inference is then given by multiplying the basic cost 15 with the factor in table 8. The weights according to this procedure can be seen in table 9. For example, Mon↓↑ gets 0.5 directionality points, 1 harmony point, 1 switch point and 1 hierarchy point, if the first quantifier is NO. That sums up to 2.5 and 3.5, respectively, yielding complexity $15 \cdot 2.5 = 38$ and $15 \cdot 3.5 = 53$, respectively for Mon↓↑-inferences (rounded to next integer). Hypothesizing that this model can be readily extended to left-side inferences, we see that their weights are:

**Table 9:** Weights for the inference rules in the QI-fragment as computed above. Numbers are rounded up.

| Inference | Mon↑↑ | Mon↑↓ | Mon↓↑ | Mon↓↓ |
|---|---|---|---|---|
| Complexity | 0/15 | 23/38 | 38/53 | 30/45 |

| Inference | ↑↑Mon | ↑↓Mon | ↓↑Mon | ↓↓Mon |
|---|---|---|---|---|
| Complexity | 0/15 | 23/38 | 23/38 | 15/30 |

Recall that they are meant to be generally easier because left-side monotonicity entailments cannot switch directions. This extension of the model will later lead to interesting predictions.

## 6 Natural Logic at Work

We think that this logic is adequate for the task of modeling the syllogistic- and the QI-fragment. Going back to our natural logic roadmap (chapter 3.2) – the inference rules presented account for all of the valid inferences in the QI-fragment, all syllogisms that are valid in Aristotelian logic or predicate calculus or both and even defines some others as good (e.g. AO3O, see table 4). Our cognitive goals imposed two constraints on the logic: firstly, apart from the logic having to account for the whole fragment, inference rules should be informed and justified by semantical reasoning or empirical results from psychology. We have extensively done this in chapter 5. Secondly, we imposed a simplicity constraint to make away with rules that are not necessary to account for the whole fragment. We have seen this in our decision not to include inferences based on smoothness above. It follows immediately, that our logic is incomplete, but completeness has never been the goal. It is even doubtful that it makes sense to apply the traditional metalogical vocabulary to NQL: both notions of soundness and completeness are relative to a fixed semantics. And while we have given semantics for generalized quantifiers above, we also noted that when it comes to their inferential properties, we prefer to remain deliberately unclear about their semantics – we prefer interpretation-independent inferences over interpretation-dependent ones as far as this is possible. We had argued further above that we prefer talk about *good* and *bad* inferences over talk about *valid* and *invalid* ones as in chapter 2, we have seen a variety of normative standards none of which has a priori priority over the others. One possible consequence of our cognitive motivation is thus to disregard the traditional metalogical vocabulary for the analysis of NQL.[15] We will now get to the last step of our roadmap and evaluate our model against our initial goals that seem more fitting than an analysis of its metalogical properties. Before we will get to this evaluation, however, we will have to make some remarks on how to interpret our results and the relationship between the complexity of a rule and mean success of reasoners in a task.

## 6.1 Complexity and Mean Success

The weights of the inference rules in our logic operationalize their respective *complexity*, which should align with the *difficulty* that reasoners experience as observed in experimental settings. The complexity of an inference from premises to conclusions is the *weighted length of proofs* as computed in appendix B for valid syllogisms. There are several interpretations of this: firstly, the weights relate the difficulty of one step to the difficulty of another, the model indicates, for example, that simple monotonicity inferences are not very hard – but still harder than conversion and pseudoconversion. Secondly, the weighted lengths of proofs relate the difficulty of one inference from premises to a conclusion to others, stating for example that the syllogism AI1I is much harder than EA1O because the latter's weighted length of proof is much shorter. Thirdly, the weighted length of proofs shows strong correlation with the mean success rates in experimental settings as given in tables 1 (syllogistic fragment) and 5 (QI-fragment). And, indeed, this will for now be the primary measure of adequacy of the weights.

This marks the point where the interpretation of the model's results become a bit difficult – the model has a *quantitative* aspect here in the sense that it can be used to

[15] Some might infer that NQL cannot be called a logic – we do however think that this is not true. This is still a systematic study of certain forms of arguments, providing a system of inferences based on semantic properties. The irritation that our stance might give raise to, we think, is due to its motivation.

predict the mean success of reasoners on an inference: the weights can be interpreted as cost that is subtracted from an initial ″cognitive reservoir“ of 100 units (as done in Geurts 2003). The values given by that not only correlate with the mean success rates in experiments but *align* with them, i.e. have a strong positive correlation (in fact, subtracting the weighted length of proof from 100 in effect only switches the algebraic sigh of the Pearson r coefficient – if we report a Person r of 0.96 for the syllogistic fragment, the weighted length of proof would give raise to a Pearson r of -0.96, without any changes to $r^2$). We will look at two examples to make this clearer.

Proof of AI3I:

$$\begin{array}{lll}
[1] & \textsc{all}(M,P) & \textit{premiss} \\
[2] & \textsc{some}(M,S) & \textit{premiss} \\
[3] & \textsc{some}(S,M) & \textit{Conv on } [2] \\
\hline
[4] & \textsc{some}(S,P) & \uparrow\!\textit{Mon on } [1] \textit{ and } [3]
\end{array}$$

Complexity = $Conv + Mon = 15$, thus predicted mean success Success = $100 - 15 = 85$.

Proof of AE2O:

$$\begin{array}{lll}
[1] & \textsc{all}(P,M) & \textit{premiss} \\
[2] & \textsc{no}(S,M) & \textit{premiss} \\
[3] & \textsc{all not}(S,M) & \textit{pConv on } [2] \\
[4] & \textsc{all not}(S,P) & \textit{Mon}\!\downarrow \textit{ on } [4] \textit{ and } [1] \\
\hline
[5] & \textsc{some not}(S,P) & \textit{exImp on } [4]
\end{array}$$

Complexity = $pConv + Mon_N + exImp \ = 95$, this predicted mean success Success = $100 - 95 = 5$.

While the step from complexity to accuracy is not entirely unproblematic and demands for a stronger commitment, we will later on make use of this to make predictions for further research. For now, we will make the step from *correlation with mean success rates* to *prediction of mean success results* – but beware of the fact that the appropriateness of this stance depends on whether the empirically testable hypotheses that can be derived from the model proof adequate.

### 6.2 The Syllogistic Fragment

Recall the complexity-hierarchy of inferences for the Syllogistic fragment that we introduced above and that is grounded in semantical and psychological considerations:

| Inference | exImp | Mon$_N$ | Mon | pConv | Conv |
|---|---|---|---|---|---|
| Complexity | 60 | 30 | 10 | 5 | 5 |

In the appendix B, we prove all syllogisms that are valid in Aristotelian logic or predicate calculus (or both) and compute the model's predictions using these weights. These proofs are intended to be minimal, we do however not have any proof that they actually are. Note that all proofs need only between one and four inferences, there is thus not much space for shortening. We will however leave this open for further research. The results, i.e. the model's predictions for valid syllogisms, can be seen in table 10.

**Table 10:** Comparison between the experimental data in Chater & Oaksford's (1999) meta-study (in brackets) and our model. Syllogisms are ordered in decreasing order of mean success. Predictions that are more than 10% apart from the actual performance are marked as gray.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| AI1I | (92) | 90 | IA3I | (85) | 90 | EA3O | (22) | 5 |
| IA4I | (91) | 85 | OA3O | (69) | 70 | AA4I | (16) | 25 |
| AA1A | (90) | 90 | AO2O | (67) | 70 | EA4O | (8) | 0 |
| AI3I | (89) | 85 | EI1O | (66) | 65 | AA1I | (5) | 30 |
| EA2E | (89) | 85 | EI2O | (52) | 60 | EA1O | (3) | 5 |
| AE2E | (88) | 90 | EI3O | (48) | 60 | EA2O | (3) | 0 |
| EA1E | (87) | 90 | AA3I | (29) | 30 | AE4O | (2) | 0 |
| AE4E | (87) | 85 | EI4O | (27) | 55 | AE2O | (1) | 5 |

A statistical analysis is, unfortunately, not very telling: Geurts' model left relatively little room for improvement. The model predicts almost all variance ($r^2 = 0.93$) and has strong correlation with performance (Pearson r = 0.96, compare that to $r^2 = 0.87$ and Pearson r = 0.93 in Geurts' original model). We wish however to point out that while, in our model, four predictions are more than 10% apart from the actual performance, in Geurts' original model, 13 predictions are (making for more than half of them – see table 2) and while in our model, two predictions are more than 20% apart from the mean success rates, in Geurts' original model, 9 prediction are.

In light of our model's stance on existential import (reasoners accept it but are reluctant to draw inferences from it), the difference between AA1A and AA1I is very telling – we will come back to these two syllogisms later.

### 6.3 The Quantifier Iteration Fragment

Our weight-assignment and the inferences that model reasoning on the syllogistic fragment is a refinement of Geurts' model but already adds a new dimension of inferences by accounting for left-side monotonicity inferences. The rules and weights on the QI-fragment mark an original contribution that has to be evaluated accordingly – we cannot compare the results of statistical analyses, as we did regarding the syllogistic fragment. The inferences in this fragment are specifically designed to investigate the impact of different combinatorial monotonicity profiles on reasoning. We will omit the proofs for this fragment as they all have length 1. The results are in table 11: we obtain $r^2 = 0.88$ and Pearson r= 0.94. NQL is thus well capable of capturing the general trends and predicts much of the variance in the empirical data. We can see that the model predicts actual cognitive difficulty on this fragment reasonably well – it is however appropriate to test the measure's generality beyond the QI-fragment: while modeling the QI-fragment was our initial goal, showing the measure's generality would show that the methodology brought forward here that combines semantic and psychological insights is indeed very fruitful.

**Table 11:** Predictions of the natural logic for the QI-fragment. Predictions that are more than 10% away from the original results are marked as gray.

| Det$_A$ | Det$_B$ | Minor | % | # | NatLog |
|---|---|---|---|---|---|
| AT LEAST↑ | MORE THAN↑ | ALL$(B,C)$ | 96 | 1 | 85 |
| EVERY↑ | MORE THAN↑ | ALL$(B,C)$ | 91 | 2 | 100 |
| MOST↑ | MORE THAN↑ | ALL$(B,C)$ | 91 | 2 | 85 |
| SOME↑ | MORE THAN↑ | ALL$(B,C)$ | 87 | 3 | 85 |
| NO↓ | FEWER THAN↓ | ALL$(B,C)$ | 73 | 4 | 70 |
| EVERY↑ | FEWER THAN↓ | ALL$(C,B)$ | 71 | 5 | 77 |
| MOST↑ | FEWER THAN↓ | ALL$(C,B)$ | 62 | 6 | 62 |
| SOME↑ | FEWER THAN↓ | ALL$(C,B)$ | 60 | 7 | 62 |
| NO↓ | MORE THAN↑ | ALL$(C,B)$ | 53 | 8 | 62 |
| AT LEAST↑ | FEWER THAN↓ | ALL$(C,B)$ | 53 | 8 | 62 |
| AT MOST↓ | MORE THAN↑ | ALL$(C,B)$ | 38 | 9 | 47 |
| AT MOST↓ | FEWER THAN↓ | ALL$(B,C)$ | 36 | 10 | 55 |

We have earlier discussed that it is our goal to connect weight-assignments to semantic properties or psychological research – the results here however indicate that other aspects have to be taken into consideration: the role of specific quantifiers and their interactions with one another. The results in Geurts and van der Silk's (2005) experiment compared with our semantically motivated weights indicate indeed a special role for AT MOST – we should however not rush no conclusions. AT MOST is semantically equivalent to FEWER THAN, it would be interesting to see whether the same effect can be observed with both quantifiers.

This might lead back to our considerations about frequency (chapter 5). Or one can come back to the considerations on cardinal quantifiers in chapter 4 which indicated that different quantifiers might have different inferential properties. We do however think that the right way to go here is to further investigate the adequacy of the model brought forward here – we will now turn to empirically testable predictions that can be derived and would shed additional light on the adequacy of the semantic, inferentialist approach of this natural logic.

# 7 Predictions

While NQL is grounded in semantic relationships and psychological evidence, it might also seem somewhat post hoc – there is barely *direct* evidence accounting for the weight-assignments but mostly *related* evidence. Luckily, the model allows for empirically testable predictions. We will now discuss them and suggest experiments.

Firstly, the natural logic gives a reason why, for example, the ″extended″ syllogism AA4A is endorsed so much more than the Aristotelian AA4I: reasoners are reluctant to draw exImp inferences because often, when premises allow for this inference, they also allow for some monotonicity inference (call this the AA-effect). And while the latter preserves information, exImp gives up an informative statement (ALL$(A, B)$) for a less informative one (SOME$(A, B)$) – reasoners have little reason to do so, if another, much more exciting and information-preserving inference can be made. We wish to test our theory's prediction that reasoners are very reluctant to draw an exImp inference if another inference preserves information. Reasoners tend to *accept* exImp but are reluctant to *make* it. The way to test this, we believe, is to actually give reasoners multiple possible conclusions for a set of premises (some of which indicate good, some of which indicate bad inferences) – reasoners have to choose the ones that follow from the premises. For example, given the syllogistic premises AA1:

NOALL$(B, C)$
ALL$(A, B)$

reasoners are presented with a set of conclusions that contains ALL$(A, C)$ and SOME$(A, C)$. Similarly, one can test the extended AA4A against the Aristotelian AA4I where the premises are ALL$(C, B)$ and ALL$(B, A)$ and the extended and Aristotelian conclusions are ALL$(C, A)$ and SOME$(A, C)$.

Secondly, our logic predicts that monotonicity inferences that involve a NOT (SOME NOT, ALL NOT) are harder than others (recall that this is our inferential operationalization of Geurts' idea that proofs involving O-propositions are harder than others). In an inferential setting, this hypothesis allows for simple testing. The data in Chater & Oaksford's (1999) meta-study seems to be in favor of this hypothesis: the two syllogisms OA3O and AO2O are precisely the only ones that only involve one application of monotonicity rules involving SOME NOT or ALL NOT – and proof to be considerably harder that those who do not. To test this hypothesis, one needs to obtain mean success rates for derivations that involve monotonicity-inferences and a NOT more than the two syllogisms OA3O and AO2O.

Thirdly, the theory predicts no impact regarding the choice of quantifiers as only monotonicity- and symmetry-properties matter. This hypothesis is very likely to be falsified – the work of Geurts & van der Silk (2005) on quantifier iteration suggests that it is very likely that downward entailing cardinal quantifiers (e.g. AT MOST) are harder to process. A more detailed investigation into this issue will however provide helpful directions for further investigations.

Finally, we wish to know if our weight assignment for reasoning on iterated quantifiers makes sense outside of the QI-fragment. The model would predict, for example, that 77% of all participants would correctly guess an inference of the form

NO↓ A played against ALL↑ B.<br>
All B were C.

---

NO↓ A played against ALL↑ C.

to be valid. This again takes the quantitative aspect of the model very serious. Such a test would however also shed light on the question whether one should actually take it so serious. As can be seen from the weight-assignments for left-side monotonicity inferences on iterated quantifiers, we can furthermore predict that they should generally be easier, as they do not involve any directionality-switch. To sum up some of the testable hypotheses that can be derived from our model:

- Reasoners accept exImp-inferences but are very reluctant to draw them themselves.
- Monotonicity inferences involving SOME NOT and ALL NOT are harder than those who do not.
- Apart from the second hypothesis, the choice of quantifier is not as important as its monotonicity properties
- Left-side monotonicity inference on iterated quantifiers are not harder than right-side monotonicity inferences, they should in fact generally be easier as they cannot involve a switch / directionality change
- The complexity-measure for the inferences operating on the QI-fragment can be extended beyond it

## 8 Interlude: Logic, Cognition, and Philosophy

We have already seen a variety of ways in which natural logic suits our goal of modeling (the cognitive difficulty of) reasoning with quantifiers. Using quantification over sets – instead first order logic quantification over individuals – and subject-predicate form, one gets a step closer towards natural language, inheriting much of its flexibility. First order logic is however not completely inadequate to deal with generalized quantifiers: van Lambalgen (1995) brings forwards an axiomatization for $Q$ as FORALMOSTALL (interpreted in a measure-theoretic sense where $Qx\phi(x)$ is interpreted as "$\{x|\phi(x)\}$ has measure 1"). While this gives proper proof-theoretic semantics for this quantifier, it does not bring the generality and flexibility that we aim for.

It is quite telling that first order logic cannot operationalize MOST, as "Proportional quantifiers have played a central role in the development of formal semantics because they set a benchmark for the expressive power needed to describe quantification in natural language" (Hackl 2009, 63). While there is a possibility of using *weak axiomatization* (see for example Mostowski 1995 and van Benthem & Westerståhl 1995) for MOST and still obtain its proof-theoretic semantics, it seems that with every step in this direction, we stray even further away from natural language. While we can confidently say that first order logic does not suit our goals – both in expressive power and requirements toward representations – we still maintain the inferential ways that were first walked by natural deduction proof systems.

We will now take a step back to provide some philosophical context for our decision to focus on inferential properties. Appendix A contains model-theoretical semantics for generalized quantifiers, and indeed, such *extensional* semantics is the predominant approach to quantifier meaning (e.g. Peters & Westerståhl 2006). There are however alternatives: we have already seen in our discussion of Gentzen's natural deduction calculus that it defines the meaning of the logical constants through introduction- and elimination-rules, i.e. through inference rules. Privileging inference over reference in the order of semantic explanation has seen some defenders in philosophy.

As is often cited in the works of philosophers and semanticists, Ludwig Wittgenstein prominently proclaimed that meaning is often the same as use:

> "For a *large* class of cases of the employment of the word 'meaning' – though not for *all* – this word can be explained in this way: the meaning of a word is its use in the language." (Wittgenstein 1953/2009, 25e)

One semantic theory that this statement has given raise to is Robert Brandom's *inferentialism* (Brandom 2001). In his theory, our understanding of a concept is shown through our correct use of it:

> "The meanings of linguistic expressions [...] should be understood, to begin with, in terms of playing a distinctive kind of role in *reasoning*." (Brandom 2001, 1)

Talk about concepts is thus talk about their distinctive role in reasoning – saying that something is so-and-so is taking a commitment to the *good* inferences that can be drawn from it. Mastering a concept thus means mastering its inferential use (Brandom 2001, 11 f.). Privileging inference over reference in the order of semantic explanation means: there are good and bad inferences in everyday life, doings that are appropriate and doings that are inappropriate. An endeavor such as a natural logic makes implicit commitments explicit, it states which inferences are appropriate and which not, but it does not define that. [16] It turns something that one can initially only do into something

[16] A similar picture is drawn by Nelson Goodman, stating that "Principles of deductive inference are justified by their conformity with accepted deductive practice" (Goodman 1983, 63). Historically speaking, then, what we call logic developed through a long phase of "mutual adjustments between rules and accepted inferences" (Goodman 1983, 64), with large parts of logic developing a life of its own, going far beyond accepted every day inferences. In this picture, logic is neither purely normative nor purely descriptive, but a little bit of both and more: logical rules inform inferential practice and vice versa, but parts of logic go far beyond inferential practice and are thus not as psychologically relevant. See also van

that one can also say. In our treatment of quantifiers, this means for example: people (as the experimental results show) are quite competent in the use of generalized quantifiers, but a natural logic can make explicit, which inferences are appropriate and which not (people do not need a semanticist to tell them what to do with SOME, but it takes one to explicitly formulate the class of appropriate inferences involved).

Thus, in inferential semantics, ″practices of giving and asking for reasons have a privileged, indeed *defining*, role“ (Brandom 2001, 14, our emphasis). This is contrary to predominant semantic theories that focus on reference, denotation, and extension.

Now, there are some problems with this theory: inferences are not everything, that people do with concepts. Large parts of any use-centric theory cannot be properly formalized. But we believe that, restricted to our topic of generalized quantifiers, we are in luck: from an inferentialist' perspective, they are relatively well-behaved. Inferentialism states that understanding a quantifier means knowing which inferences it allows for. A natural logic as presented above can now make explicit what to do with quantifiers (though it does not define them uniquely, it rather speaks of classes of quantifiers that have certain monotonicity or symmetry properties). However, it cannot constitute a full-fledged semantic theory – the logic presented above does not distinguish between non-symmetric quantifiers that have the same monotonicity profile. This discussion also allows us to draw a formal line between natural deduction and natural logics (apart from the fact that the latter should operate on natural language surface): while the former fully specified the semantics of all logical constants relevant to it, a natural logic usually does not.

Inferentialism tries to reflect the shift occurring in the twentieth century towards the explanatory priority of what humans do with language (e.g. Austin, Sellars, Wittgenstein). In the end, we do however have to say that extensional and intensional semantics do not differ much from a cognitive scientists' perspective: we initially approached quantifiers traveling on extensional roads, which led to the same results. But inferentialism allows for a fresh perspective on how these results are situated in human practice. It aims to make reasoners' implicit commitments explicit in formalizing reasoning patterns that are appropriate. The role of an inference rule is thus to make explicit, what before was implicit in human practice.

The inferentialist approach thus defines what kind of inference is *good* and *bad* and how hard the good ones are. It also informs about the essential semantic properties of expressions, as those are the ones that we derive inferences from, the ones we put to *use*. We have however also seen the limits of the inferentialist perspective: there are very common, meaningful quantifier expressions that are low on inferential properties, such as FEW.

Benthem (2008)

# 9 Finale

We have now spent a considerable amount of time at the intersection of logic, linguistics and psychology. We have thoroughly motivated our choice of modeling approach – this motivation was mostly connected to the cognitive plausibility that comes with natural language representations and the flexibility that a complexity measure grounded in a weighted length of proof allows for. We can directly note the two main problems that we encountered: firstly, there is almost only related empirical evidence for weight-assignments, almost no direct evidence. There is thus some post-hoc'ness in the air. Secondly, we have encountered some problems in accounting for all of the data with our approach – by taking only a limited amount of semantic properties into account, we cannot explain the differences between the cognitive difficulty of inferences using AT MOST and inferences using NO (table 11).

That being said, there are obvious directions for further research. Firstly, we derived empirically testable hypotheses from NQL and suggested experiments as to test them. This is allowed by our use of informative weight assignments, which can thereby tested for their adequacy.

Secondly, we saw limitations in the combinatorial approach to the monotonicity profiles of iterated quantifiers – to the best of our knowledge, there are no mathematical results that allow for a definition of combinatorial monotonicity profiles based solely on the monotonicity profiles of the single quantifiers outside of the QI-fragment. Such a result would however greatly expand natural language fragment modeled by NQL.

Thirdly, weights could be further optimized as to increase the fit with the empirical data. We do however not think that this has priority: the statistical reference numbers are already considerably high and the ratio of datapoints to parameters invites overfitting. The important fact of the weights is that they are informative and that they put the complexity of different inferences in relation to one another.

We have however seen that the connection of semantic and psychological results can provide a powerful tool for the derivation of hypotheses. The natural logic approach relies heavily on the fact that some natural language expressions show strong inferential properties that can be put to use. NQL successfully models reasoning on the syllogistic and QI-fragments, explaining much of the variance in mean success rates. The weights are informative and well-grounded in psychological research, allowing for clear directions for further research. The fact that NQL managed to better capture the large variety of cognitive difficulties in syllogistic reasoning much better than competing approaches (recall table 4) invites the use of this approach to other realms of reasoning – we are convinced that, as evidence stands now, this is the right hammer for this kind of nail.

# Appendices

## A Generalized Quantifier Theory

We will give an overview over the model-theoretic semantics that lie at the heart of generalized quantifier theory. For our presentation of generalized quantifier theory and the quantifier-specific notation that we use throughout the book, we will largely follow Peters & Westerståhl (2006).

There is however good reason to limit our investigation to quantifier expressions that syntactically are either determiners in noun phrases ("MOST clowns are creepy") or noun phrases themselves ("EVERYONE is creepy"): these kinds of quantifiers have been heavily studied and are also the ones that are present in Geurts & van der Silk's (2005) dataset that we derive our cognitive motivation from. The understanding of quantifier expressions as determiners in NPs and of quantifier expressions as NPs will correspond to the semantic distinction between type <1,1> and type <1> quantifiers that will be introduced and used subsequently (Peters & Westerståhl 2006, 11).

Furthermore, the logical quantifiers $\exists$ and $\forall$ quantify over individuals (i.e. using count nouns like "lamp" or "book"), thereby ignoring quantification over collective count nouns ("crowd") or mass nouns ("furniture", Peters & Westerståhl 2006, 1). Similarily, as we will see later, we will only allow for first order quantification (over individuals), as opposed to second order quantification (e.g. over relations between individuals as well).

### A.1 Definitions: of Models, Truth and Quantifiers

We will need to have a language (syntax), a class of models or interpretations and to fix the truth relation between sentences in the language and the models (we will explicitly follow Peters & Westerståhl 2006).

**Vocabularies and Models**
A (first-order) *vocabulary* is a set $V$ of non-logical symbols: individual constants, predicate symbols (of various arities), and function symbols (also of various arities). $V$ is allowed to be empty. It is *relational* if it has only predicate symbols. A *model* (for the vocabulary $V$) has the form

$$\mathcal{M} = (M, I)$$

where $M$ is a (usually non-empty) set – the universe – and $I$ is an interpretation function, which assigns a suitable interpretation $I(u)$ to each item $u \in V$ where $I(c) \in M$ if $c$ is an individual constant; $I(P) \subseteq M^n$ if $P$ is a $n$-ary predicate symbol, etc. We will assume throughout this work that our vocabularies are relational.

We assume a certain familiarity with the *logical symbols* $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\Leftrightarrow$, $=$ and the existential and universal quantifier, $\exists$ and $\forall$, respectively. As mentioned, the *non-logical* *symbols* are the ones appearing in our vocabulary $V$. Parentheses are used according to the usual conventions. We now turn to the definition of formulas and sentences.

**Formulas and Sentences**
Let $V$ be a relational vocabulary, i.e. only having predicate-, but no function-symbols. The *V-formulas* are defined inductively as follows:

i If $P$ is a $n$-ary predicate symbol and $x_1, x_2, ..., x_n$ are variables, then $P(x_1, x_2, ..., x_n)$ is a $V$-formula.

ii If $x$ and $y$ are variables, the $(x = y)$ is a $V$-formula.

iii If $\phi$ and $\psi$ are $V$-formulas, then so are $\neg\phi$, $(\phi \wedge \psi)$, $(\phi \vee \psi)$, $(\phi \rightarrow \psi)$ and $(\phi \leftrightarrow \psi)$.

iv If $\phi$ is a $V$-formula and $x$ a variable, then $\forall x\phi$ and $\exists x\phi$ are $V$-formulas.

Nothing else is a $V$-formula. A *V-sentence* is a $V$-formula without any free variables.

We can now turn to the model theoretic semantics, hinging on the satisfaction relation

$$\mathcal{M} \vDash \phi(a_1, ..., a_n)$$

saying that $\phi(a_1, ..., a_n)$ is *true* in $\mathcal{M}$, where $\phi(x_1, ..., x_n)$ is a $V$-formula and $a_1, ..., a_n$ is assigned to the free variables $x_1, ..., x_n$. If $\phi$ is a $V$-sentence, i.e. without free variables, then we say

$$\mathcal{M} \vDash \phi$$

Still assuming that our vocabularies are relational, the truth definition of first order logic (FOL) goes as follows:

**FOL truth definition** following Peters & Westerståhl (2006).

i $\mathcal{M} \vDash P(a_1, ..., a_n)$ iff $(a_1, ..., a_n) \in I(P)$ when $P(x_1, ..., x_n)$ is an atomic formula.

ii $\mathcal{M} \vDash a_i = a_j$ iff $a_i$ is the same member of $M$ as $a_j$ for an atomic formula $x_i = x_j$.

iii $\mathcal{M} \vDash \neg\phi(a_1, ..., a_n)$ iff $\mathcal{M} \nvDash \phi(a_1, ..., a_n)$

iv $\mathcal{M} \vDash (\phi \wedge \psi)(a_1, ..., a_n)$ iff $\mathcal{M} \vDash \phi(a_1, ..., a_n)$ and $\mathcal{M} \vDash \psi(a_1, ..., a_n)$ (and analogous for the other connectives).

v $\mathcal{M} \vDash \forall x\phi(x, a_1, ..., a_n)$ iff for all $b \in M$, $\mathcal{M} \vDash \phi(b, a_1, ..., a_n)$.

vi $\mathcal{M} \vDash \exists x\phi(x, a_1, ..., a_n)$ iff for some $b \in M$, $\mathcal{M} \vDash \phi(b, a_1, ..., a_n)$.

Note that these truth definitions can only be understood by someone who already understands English, i.e. who can make intelligible what *and*, *all* and *some* mean (Peters & Westerståhl 2006, 58). We will revisit this point later on. Before we introduce a formal treatment of generalized quantifiers, we need one more definition.

**Extension of a Formula in a Model**
Given a formula $\phi = \phi(x, y_1, ..., y_n) = \phi(x, \bar{y})$ and a ordered set $\bar{a}$ of $n$ objects in $M$,

$$\phi(x, \bar{a})^{\mathcal{M},x} = \{b \in M | \mathcal{M} \vDash \phi(b, \bar{a})\}$$

is called the *extension* of $\phi$ in $\mathcal{M}$.

For example, if one takes the familiar quantifiers of first order logic,

$$\mathcal{M} \vDash \forall x \phi(x, \bar{a})^{\mathcal{M},x} \text{ iff } \phi(x, \bar{a})^{\mathcal{M},x} = M$$

$$\mathcal{M} \vDash \exists x \phi(x, \bar{a})^{\mathcal{M},x} \text{ iff } \phi(x, \bar{a})^{\mathcal{M},x} \neq \emptyset$$

This notation denoting the extension of a formula allows us to define a quantifiers truth conditions in set-theoretic terms and will be used heavily in what is to follow. If the context is sufficiently clear, we will however sometimes leave out the superscripts and write $\phi(x, \bar{a})$ instead of $\phi(x, \bar{a})^{\mathcal{M},x}$. Defining the semantics of generalized quantifiers through their respective extensions is not our final goal but provides us with a helpful intuition to later find suitable axioms and inference rules for our inferential (intensional) semantics. Given this background, we can finally define generalized quantifiers. We will thereby distinguish two semantic types. While there are more types, the ones we introduce here are the ones relevant to our investigation and also the ones that are most studied.

## A.2 Type <1>: Noun Phrases

Type <1> quantifiers are a curious construction: while they are predominant in first order logic, there are natural languages that do not even have any means of expressing type <1> quantification.

**Type <1> Quantifiers**
For a universe $M$, we let $Q_M$ be any set of subsets of $M$, and use at the same time (to simplify notation) '$Q$' as a new symbol as a variable-binding operator. Then $Q$ is a *generalized quantifier of type* <1>, whose meaning is given by

$$\mathcal{M} \vDash Qx\phi(x, \bar{a}) \text{ iff } \phi(x, \bar{a})^{\mathcal{M},x} \in Q_M$$

The two quantifiers of first order logic, the familiar universal and existential quantifiers, $\forall$ and $\exists$, respectively, have this form. While these present the logically simplest case, they are not dominant in natural language. Lets look at them.

(i) $\exists x \phi(x)$ – SOMETHING is $\phi$

(ii) $\forall x \phi(x)$ – EVERYTHING is $\phi$

Quantification here is over the whole universe $M$. Note that those are different from

(i) $\text{ALL}(A, B)$ – All $A$s are $B$s

(ii) $\text{SOME}(A, B)$ –

Those quantifiers are restricted to the set $A \subseteq M$. We can however still express them using first order logic:

(i) $\forall x(A(x) \to B(x))$ – Everything that is an $A$ is also a $B$.

(ii) $\exists x(A(x) \wedge B(x))$ – There is something that is an $A$ and also a $B$.

One can even formalize cardinal quantifiers

(i) AT LEAST $3(A, B)$

(ii) $\exists x \exists y \exists z(x \neq y \wedge x \neq z \wedge y \neq z \wedge A(x) \wedge B(x) \wedge A(y) \wedge B(y) \wedge A(z) \wedge B(z))$

With cardinal quantifiers, the limits of first order quantification become obvious. While they may be logically equivalent, they are linguistically quite different. We will however not spend too much time with type <1> quantifiers – they will return from time to time but for now, it is sufficient to note that while they can formalize the sentences above logically properly, they do not provide linguistically equivalent sentences. Furthermore, they are limited to quantification over individuals.

## A.3 Type <1,1>: Determiners

While type <1> quantifiers are the only type used in first order logic, type <1,1> quantifiers are actually predominant in the English language. With them, we can treat binary relations between sets of stuff (Peters & Westerståhl 2006, 11).

**Type <1,1> Quantifiers**
A *generalized quantifier of type* <1,1> associates with each universe $M$ a binary relation $Q_M$ between subsets of $M$. Using the same symbol as a variable-binding operator, the meaning of a quantified formula $Qx(\phi, \psi)$, where $\phi = \phi(x, x_1, ..., x_n$ and $\psi = \psi(x, x_1, ..., x_n)$ have at most the free variables shown, is given by

$$\mathcal{M} \models Qx(\phi(x, \bar{a}), \psi(x, \bar{a})) \text{ iff } Q(\phi(x, \bar{a})^{\mathcal{M},x}, \psi(x, \bar{a})^{\mathcal{M},x})$$

Characterizing type <1,1> quantifiers like this makes clear that they are a restriction of type <1> quantifiers in the following way: the first argument identifies a subset of the universe $M$ as the relevant domain of quantification (and is thus called the *restriction*), while the second argument provides its *scope*. In the example "MOST semanticists are linguists", the first argument, *semanticists*, restrict the domain of quantification to all semanticists, while *linguists* identifies the scope. Let us consider some examples.

(i) MOST(semanticists, linguists)

(ii) SOME(philosophers, semanticists)

(iii) ALL(philosophers, successful)

(iv) NO(linguists, philosophers)

(v) AT MOST 3(semanticists, psychologists)

We can note immediately that the determiner ALL shows some similarities to the type <1> logical quantifier $\forall$, and as we noted further above, we can actually express ALL(philosophers, successful) using $\forall$ as well. However, the determiner ALL requires existential import, while the logical quantifier $\forall$ does not. As a consequence of this, $\forall x(philosophers(x) \rightarrow successful(x))$ would be true if there was no philosopher, while ALL(philosophers, successful) would not.

Recalling the definition of a type <1,1> quantifier above, we can observe that the right hand side of the definition states that two sets are in a relation $Q$ with another – and the relations are as we defined them in chapter 4. Note however that this is in principle not the limit of the concept: quantifiers can also be relations between relations, i.e. second order relations.

## A.4 Monotonicity Revisited: Smoothness

To fully capture the monotonicity behavior of MOST's first position, we have to make a quick digression. Monotonicity behavior can be illustrated in *number triangles*, which have the following form.

The use of number triangles to further characterize generalized quantifiers was pointed out by van Benthem (1986) and subsequently advocated by Strössler (2017) and Peters & Westerståhl (2006). Number triangles will point us to the fact that classifying monotonicity properties on a yes-or-no scale is too coarse – especially in the case of left-side behavior, one is in need of a more detailed classification to obtain all the monotonicity profiles necessary to get hold of a proper natural logic based on monotonicity properties. The idea behind number triangles is the following: some determiners can be described using two numbers (Strößler 2017).

**Number Triangles for Quantifiers**
Let $Q(A, B)$ be a quantifier with arguments $A$ and $B$, $m = |A \cap B|$ and $k = |A - B|$ (note that $|A| = m + k$). The number triangle is given by a set of tuples formed as

follows.

$$\begin{array}{ccccccccc}
 & & & & 0,0 & & & & \\
 & & & 0,1 & & 1,0 & & & \\
 & & 0,2 & & 1,1 & & 2,0 & & \\
 & 0,3 & & 1,2 & & 2,1 & & 3,0 & \\
0,4 & & 1,3 & & 2,2 & & 3,1 & & 4,0 \\
\multicolumn{9}{c}{\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots}
\end{array}$$

Where the $i$-th row represents the situation for $|A| = i - 1$, the first row thus represents the situation for $|A| = 0$ and the left hand side is $m$ and the right hand side number is $k$. Such a tree represents all possible relationships between $m$ and $k$ for all cardinalities of $A$.
The *number triangle for a quantifier* $Q$ is now a corresponding triangle where we set a ”+“ for all the relationships between $m$ and $k$ that make $Q(A, B)$ true and a ”−“ if otherwise. Take for example the number triangle for MOST:

$$\begin{array}{ccccccccc}
 & & & & - & & & & \\
 & & & - & & + & & & \\
 & & - & & - & & + & & \\
 & - & & - & & + & & + & \\
- & & - & & - & & + & & + \\
\multicolumn{9}{c}{\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots}
\end{array}$$

And the number triangle for NO:

$$\begin{array}{ccccccccc}
 & & & & - & & & & \\
 & & & + & & - & & & \\
 & & + & & - & & - & & \\
 & + & & - & & - & & - & \\
+ & & - & & - & & - & & - \\
\multicolumn{9}{c}{\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots}
\end{array}$$

Other authors have decided to leave the uppermost entry undefined (e.g. Strößler 2017), but we decided to define it as ”-“ to emphasize our decisions regarding existential import. Peters and Westerståhl (2006, 177 ff.) describe the relationship between the number triangle for a quantifier $Q$ and its monotonicity properties in detail.

If one was standing on a ”+“-position in the number triangle for any quantifier

$Q(A,B)$, the question is, in which directions one can ”walk“ without reaching a ”-“-position. As for the example of NO above, the two safe directions would be southwest (SW) and northeast (NE). We will now go on to define four different left-side monotonicity profiles for quantifiers $Q(A,B)$. Definitions for type <1> quantifiers go analogously, but we will leave them out here.

(i) SE-Mon

$$Q(A,B) \wedge A \subseteq A' \wedge A - B = A' - B \Rightarrow Q(A',B)$$

(ii) SW-Mon

$$Q(A,B) \wedge A \subseteq A' \wedge A \cap B = A' \cap B \Rightarrow Q(A',B)$$

(iii) NW-Mon

$$Q(A,B) \wedge A' \subseteq A \wedge A - B = A' - B \Rightarrow Q(A',B)$$

(iv) NE-Mon

$$Q(A,B) \wedge A' \subseteq A \wedge A \cap B = A' \cap B \Rightarrow Q(A',B)$$

As we see here, this finer distinction between left-side monotonicity profiles allows for more inferences. We can now define an additional kind of left-side monotonicity behavior.

**Smoothness** A quantifier $Q$ is *smooth* iff it is NE-Mon and SE-Mon, i.e. the following two conditions hold:

(i) $Q(A,B) \wedge A \subseteq A' \wedge A - B = A' - B \Rightarrow Q(A',B)$

(ii) $Q(A,B) \wedge A' \subseteq A \wedge A \cap B = A' \cap B \Rightarrow Q(A',B)$

This gives us a better hold on the proportional quantifiers: they are all smooth. We will shortly use (or rather: explain why we will not use) this concept in chapter 5.

# B Syllogisms: Natural Logic Proofs and Complexity Calculations

Recall the weights used for inference rules.

| Inference Rule | exImp | $\text{Mon}_N$ | Mon | pConv | Conv |
|---|---|---|---|---|---|
| Weight | 60 | 30 | 10 | 5 | 5 |

We will look at the proofs and associated complexity computations. The proofs are ordered in terms of success in Chater & Oaksford’s (1999) meta-study (see table 1) and contain all 24 syllogisms that are valid in Aristotelian or predicate logic, or both. It does however *not* refer to the 256 additional extended syllogisms as there is no sufficient empirical data available. The proofs are ordered in terms of success in the syllogism meta-study. Here is AI1I:

$$\begin{array}{lll} [1] & \textsc{all}(M,P) & \textit{premiss} \\ [2] & \textsc{some}(S,M) & \textit{premiss} \\ \hline [3] & \textsc{some}(S,P) & \textit{Mon}\!\uparrow \textit{ on } [1] \textit{ and } [3] \end{array}$$

Model prediction: $100 - \textit{Mon} = 90$

Proof of IA4I:

$$\begin{array}{lll} [1] & \textsc{some}(P,M) & \textit{premiss} \\ [2] & \textsc{all}(M,S) & \textit{premiss} \\ [3] & \textsc{some}(M,P) & \textit{Conv on } [1] \\ \hline [4] & \textsc{some}(S,P) & \uparrow\!\textit{Mon on } [2] \textit{ and } [3] \end{array}$$

Model prediction: $100 - \textit{Conv} - \textit{Mon} = 85$

Proof of AA1A:

$$\begin{array}{lll} [1] & \textsc{all}(M,P) & \textit{premiss} \\ [2] & \textsc{all}(S,M) & \textit{premiss} \\ \hline [3] & \textsc{all}(S,P) & \textit{Mon}\!\uparrow \textit{ on } [1] \ \&\ [2] \end{array}$$

Model prediction: $100 - \textit{Mon} = 90$

Proof of AI3I:

$$\begin{array}{lll} [1] & \textsc{all}(M,P) & \textit{premiss} \\ [2] & \textsc{some}(M,S) & \textit{premiss} \\ [3] & \textsc{some}(S,M) & \textit{Conv on } [2] \\ \hline [4] & \textsc{some}(S,P) & \uparrow\!\textit{Mon on } [1] \textit{ and } [3] \end{array}$$

Model prediction: $100 - \textit{Conv} - \textit{Mon} = 85$

Proof of EA2E:

$$\begin{array}{lll} [1] & \textsc{no}(P,M) & \textit{premiss} \\ [2] & \textsc{all}(S,M) & \textit{premiss} \\ [1] & \textsc{no}(M,P) & \textit{Conv on } [1] \\ \hline [3] & \textsc{no}(S,P) & \textit{Mon}\!\downarrow \textit{ on } [2] \ \&\ [3] \end{array}$$

Model prediction: $100 - \textit{Conv} - \textit{Mon} = 85$

Proof of AE2E:

$$\begin{array}{lll} [1] & \textsc{all}(P,M) & \textit{premiss} \\ [2] & \textsc{no}(S,M) & \textit{premiss} \\ \hline [3] & \textsc{no}(S,P) & \downarrow\!\textit{Mon on } [1] \ \& \ [2] \end{array}$$

Model prediction: $100 - \mathit{Mon} = 90$

Proof of EA1E:

$$\begin{array}{lll} [1] & \textsc{no}(M,P) & \textit{premiss} \\ [2] & \textsc{all}(S,M) & \textit{premiss} \\ \hline [3] & \textsc{no}(S,P) & \downarrow\!\textit{Mon on } [1] \ \& \ [2] \end{array}$$

Model prediction: $100 - \mathit{Mon} = 90$

Proof of AE4E:

$$\begin{array}{lll} [1] & \textsc{all}(P,M) & \textit{premiss} \\ [2] & \textsc{No}(M,S) & \textit{premiss} \\ [2] & \textsc{No}(S,M) & \textit{Conv on } [2] \\ \hline [3] & \textsc{No}(S,P) & \textit{Mon}\!\downarrow \textit{on } [1] \textit{ and } [3] \end{array}$$

Model prediction: $100 - \mathit{Conv} - \mathit{Mon} = 85$

Proof of IA3I:

$$\begin{array}{lll} [1] & \textsc{some}(M,P) & \textit{premiss} \\ [2] & \textsc{all}(M,S) & \textit{premiss} \\ \hline [3] & \textsc{some}(S,P) & \uparrow\!\textit{Mon on } [1] \textit{ and } [2] \end{array}$$

Model prediction: $100 - \mathit{Mon} = 90$

Proof of OA3O:

$$\begin{array}{lll} [1] & \textsc{some not}(M,P) & \textit{premiss} \\ [2] & \textsc{all}(M,S) & \textit{premiss} \\ \hline [3] & \textsc{some not}(S,P) & \uparrow\!\textit{Mon on } [1] \textit{ and } [2] \end{array}$$

Model prediction: $100 - \mathit{Mon}_N = 70$

Proof of AO2O:

$$\begin{array}{lll} [1] & \textsc{all}(P,M) & \textit{premiss} \\ [2] & \textsc{some not}(S,M) & \textit{premiss} \\ \hline [3] & \textsc{some not}(S,P) & \textit{Mon}\!\downarrow \textit{on } [1] \textit{ and } [2] \end{array}$$

Model prediction: $100 - Mon_N = 70$. Note that NOT changes directionality.

Proof of EI1O:

$$\begin{array}{lll}
[1] & \textsc{no}(M,P) & \textit{premiss} \\
[2] & \textsc{some}(S,M) & \textit{premiss} \\
[3] & \textsc{all not}(M,P) & \textit{pConv on } [\textit{1}\ ] \\
\hline
[4] & \textsc{some not}(S,P) & \textit{Mon}\uparrow \textit{ on } [2] \textit{ and } [3]
\end{array}$$

Model prediction: $100 - pConv - Mon_N = 65$

Proof of EI2O:

$$\begin{array}{lll}
[1] & \textsc{no}(P,M) & \textit{premiss} \\
[2] & \textsc{some}(S,M) & \textit{premiss} \\
[3] & \textsc{no}(M,P) & \textit{Conv on } [1] \\
[4] & \textsc{all not}(M,P) & \textit{pConv on } [\textit{3}\ ] \\
\hline
[5] & \textsc{some not}(S,P) & \textit{Mon}\uparrow \textit{ on } [2] \textit{ and } [4]
\end{array}$$

Model prediction: $100 - Conv - pConv - Mon_N = 60$

Proof of EI3O:

$$\begin{array}{lll}
[1] & \textsc{no}(M,P) & \textit{premiss} \\
[2] & \textsc{some}(M,S) & \textit{premiss} \\
[3] & \textsc{some}(S,M) & \textit{Conv on } [2] \\
[4] & \textsc{all not}(M,P) & \textit{pConv on } [1] \\
\hline
[5] & \textsc{some not}(S,P) & \textit{Mon}\uparrow \textit{ on } [3] \textit{ and } [4]
\end{array}$$

Model prediction: $100 - Conv - pConv - Mon_N = 60$

Proof of AA3I:

$$\begin{array}{lll}
[1] & \textsc{all}(M,P) & \textit{premiss} \\
[2] & \textsc{all}(M,S) & \textit{premiss} \\
[3] & \textsc{some}(M,P) & \textit{exImp on } [1] \\
\hline
[4] & \textsc{some}(S,P) & \uparrow\textit{Mon on } [2]\ \&\ [3]
\end{array}$$

Model prediction: $100 - exImp - Mon = 30$

Proof of EI4O:

$$\begin{array}{lll}
[1] & \textsc{no}(P,M) & \textit{premiss} \\
[2] & \textsc{some}(M,S) & \textit{premiss} \\
[3] & \textsc{some}(S,M) & \textit{Conv on } [2] \\
[4] & \textsc{no}(P,M) & \textit{Conv on } [1] \\
[5] & \textsc{all not}(P,M) & \textit{pConv on } [4] \\
\hline
[6] & \textsc{some not}(S,P) & \textit{Mon}\uparrow \textit{ on } [2] \textit{ and } [5]
\end{array}$$

Model prediction: $100 - Conv - Conv - pConv - Mon_N = 55$

Proof of EA3O:

$$\begin{array}{lll}
[1] & \textsc{no}(M,P) & \textit{premiss} \\
[2] & \textsc{all}(M,S) & \textit{premiss} \\
[3] & \textsc{all not}(M,P) & \textit{pConv on } [1] \\
[4] & \textsc{some not}(M,P) & \textit{exImp on } [3] \\
\hline
[5] & \textsc{some not}(S,P) & \uparrow\textit{Mon on } [2] \textit{ and } [4]
\end{array}$$

Model prediction: $100 - pConv - exImp - Mon_N = 5$

Proof of AA4I:

$$\begin{array}{lll}
[1] & \textsc{all}(P,M) & \textit{premiss} \\
[2] & \textsc{all}(M,S) & \textit{premiss} \\
[3] & \textsc{some}(P,M) & \textit{exImp on } [1] \\
[4] & \textsc{some}(M,P) & \textit{Conv on } [3] \\
\hline
[5] & \textsc{some}(S,P) & \uparrow\textit{Mon on } [2] \ \&\ [4]
\end{array}$$

Model prediction: $100 - exImp - Conv - Mon = 25$

Proof of EA4O:

$$\begin{array}{lll}
[1] & \textsc{no}(P,M) & \textit{premiss} \\
[2] & \textsc{all}(M,S) & \textit{premiss} \\
[3] & \textsc{no}(M,P) & \textit{Conv on } [1] \\
[4] & \textsc{all not}(M,P) & \textit{pConv on } [3] \\
[5] & \textsc{some not}(M,P) & \textit{exImp on } [4] \\
\hline
[6] & \textsc{some not}(S,P) & \uparrow\textit{Mon on } [2] \textit{ and } [5]
\end{array}$$

Model prediction: $100 - Conv - pConv - exImp - Mon_N = 0$

Proof of AA1I:

$$\begin{array}{lll}
[1] & \textsc{all}(M,P) & \textit{premiss} \\
[2] & \textsc{all}(S,M) & \textit{premiss} \\
[3] & \textsc{some}(S,M) & \textit{exImp on } [2] \\
\hline
[4] & \textsc{some}(S,P) & \textit{Mon}\uparrow \textit{ on } [1] \textit{ and } [3]
\end{array}$$

Model prediction: $100 - exImp - Mon = 30$

Proof of EA1O:

$$\begin{array}{lll}
[1] & \textsc{no}(M,P) & \textit{premiss} \\
[2] & \textsc{all}(S,M) & \textit{premiss} \\
[3] & \textsc{all not}(M,P) & \textit{pConv on } [1] \\
[4] & \textsc{all not}(S,P) & \downarrow\textit{Mon on } [2] \textit{ and } [3] \\
\hline
[5] & \textsc{some not}(S,P) & \textit{exImp on } [4]
\end{array}$$

Model prediction: $100 - pConv - Mon_N - exImp = 5$

Proof of EA2O:

$$\begin{array}{lll}
[1] & \textsc{no}(P,M) & \textit{premiss} \\
[2] & \textsc{all}(S,M) & \textit{premiss} \\
[3] & \textsc{no}(M,P) & \textit{Conv on } [1] \\
[4] & \textsc{all not}(M,P) & \textit{pConv on } [3] \\
[5] & \textsc{all not}(S,P) & \downarrow\textit{Mon on } [4] \textit{ and } [2] \\
\hline
[6] & \textsc{some not}(S,P) & \textit{exImp on } [5]
\end{array}$$

Model prediction: $100 - Conv - pConv - Mon_N - exImp = 0$

Proof of AE4O:

$$\begin{array}{lll}
[1] & \textsc{all}(P,M) & \textit{premiss} \\
[2] & \textsc{no}(M,S) & \textit{premiss} \\
[3] & \textsc{no}(S,M) & \textit{Conv on } [2] \\
[4] & \textsc{all not}(S,M) & \textit{pConv on } [3] \\
[5] & \textsc{all not}(S,P) & \textit{Mon}\downarrow \textit{ on } [4] \textit{ and } [1] \\
\hline
[6] & \textsc{some not}(S,P) & \textit{exImp on } [5]
\end{array}$$

Model prediction: $100 - Conv - pConv - Mon_N - exImp = 0$

Proof of AE2O:

$$\begin{array}{lll}
[1] & \textsc{all}(P,M) & \textit{premiss} \\
[2] & \textsc{no}(S,M) & \textit{premiss} \\
[3] & \textsc{all not}(S,M) & \textit{pConv on } [2] \\
[4] & \textsc{all not}(S,P) & \textit{Mon}{\downarrow}\textit{ on } [4] \textit{ and } [1] \\
\hline
[5] & \textsc{some not}(S,P) & \textit{exImp on } [4]
\end{array}$$

Model prediction: $100 - pConv - Mon_N - exImp = 5$

## B.1 Additional Syllogisms in Geurts' System

Geurts (2003) provided no data on the following syllogisms – but they are valid in their system and we use the following proofs to supply the numbers with an asterisk in table 4 to complete it. Proof of IE4O:

$$\begin{array}{lll}
[1] & \textsc{some}(P,M) & \textit{premiss} \\
[2] & \textsc{no}(M,S) & \textit{premiss} \\
[3] & \textsc{all not}(M,S) & \textit{pConv on } [2] \\
[4] & \textsc{some not}(P,S) & \textit{Mon}{\uparrow}\textit{ on } [1] \ \& \ [3] \\
\hline
[5] & \textsc{some not}(S,P) & \textit{Conv on } [4]
\end{array}$$

Proof of IE3O:

$$\begin{array}{lll}
[1] & \textsc{some}(M,P) & \textit{premiss} \\
[2] & \textsc{no}(M,S) & \textit{premiss} \\
[3] & \textsc{some}(P,M) & \textit{Conv on } [1] \\
[4] & \textsc{all not}(M,S) & \textit{pConv on } [2] \\
[5] & \textsc{some not}(P,S) & \textit{Mon}{\uparrow}\textit{ on } [3] \ \& \ [4] \\
\hline
[6] & \textsc{some not}(S,P) & \textit{Conv on } [5]
\end{array}$$

Proof of IE2O:

$$\begin{array}{lll}
[1] & \textsc{some}(P,M) & \textit{premiss} \\
[2] & \textsc{no}(S,M) & \textit{premiss} \\
[3] & \textsc{no}(M,S) & \textit{Conv on } [2] \\
[4] & \textsc{all not}(M,S) & \textit{pConv on } [3] \\
[5] & \textsc{some not}(P,S) & \textit{Mon}{\uparrow}\textit{ on } [1] \ \& \ [4] \\
\hline
[6] & \textsc{some not}(S,P) & \textit{Conv on } [5]
\end{array}$$

Proof of IE1O:

$$
\begin{array}{lll}
[1] & \textsc{some}(M,P) & \textit{premiss} \\
[2] & \textsc{no}(S,M) & \textit{premiss} \\
[3] & \textsc{some}(P,M) & \textit{Conv on } [1] \\
[4] & \textsc{no}(M,S) & \textit{Conv on } [2] \\
[5] & \textsc{all not}(M,S) & \textit{pConv on } [4] \\
[6] & \textsc{some not}(P,S) & \textit{Mon}\uparrow \textit{ on } [3] \ \textit{\&} \ [5] \\
\hline
[7] & \textsc{some not}(S,P) & \textit{Conv on } [6]
\end{array}
$$

Proof of AO3O:

$$
\begin{array}{lll}
[1] & \textsc{all}(M,P) & \textit{premiss} \\
[2] & \textsc{some not}(M,S) & \textit{premiss} \\
[3] & \textsc{some not}(P,S) & \uparrow\textit{Mon on } [1] \ \textit{\&} \ [2] \\
\hline
[6] & \textsc{some not}(S,P) & \textit{Conv on } [3]
\end{array}
$$

# References

1. Barwise, J. and Cooper, R. (1981). Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*, Vol. 4, No.2: 159–219.

2. Begg, I. and Denny, J.P. (1969). Empirical reconciliation of atmosphere and conversion interpretations of syllogistic reasoning. *Journal of Experimental Psychology*, 81: 351–354.

3. van Benthem, Johan (1986). *Essays in Logical Semantics*. D. Reidel Publishing Company, Doordrecht, Holland.

4. van Benthem, Johan and Westerståhl, Dag (1995). Directions in Generalized Quantifier Theory. *Studia Logica*, 55: 389–419.

5. van Benthem, Johan (2007). A Brief History of Natural Logic. *Available Online* http://www.illc.uva.nl/Research/Publications/Reports/PP-2008-05.text.pdf

6. van Benthem, Johan (2008). Logic and Reasoning: do the facts matter?. *Studia Logica*, 88: 67–84.

7. Besold, Ta7rek R. and d'Avila Garcez, Artur and Stenning, Keith and van der Torre, Leendert and van Lambalgen, Michiel (2017). Reasoning in Non-Probabilistic Uncertainty: Logic Programming and Neural-Symbolic Computing as Examples. *Minds & Machines*, 27: 37. doi:10.1007/s11023-017-9428-3

8. Braine, Martin D.S. (1978). On the Relation Between the Natural Logic of Reasoning and Standard Logic. *Psychological Review*, Vol. 85, Nr. 1: 1–21.

9. Braine, Martin D.S. (1990). The "Natural Logic" Approach to Reasoning. *Reasoning, Necessity, and Logic: Developmental Perspectives*, Willis F. Overton (ed.): 133–157.

10. Brandom, Robert (2001). Articulating Reasons - An Introduction to Inferentialism. Harvard University Press, Cambridge, Massachusetts.

11. Chater, N. and Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, 22: 97–117.

12. Clark, H. H. (1973). Space, time, semantics, and the child. *Cognitive Development and the Acquisition of Language* (T. Moore (ed.)), Academic Press, New York: 27–63.

13. Clark, H. H. (1974). Semantics and comprehension. *Current Trends in Linguistics, Volume 12*, Mouten, The Hague: 1291–1428.

14. Endrullis, J. and Moss, L.M. (2015) Syllogistic Logic with "Most". *International Workshop on Logic, Language, Information, and Computation* (V. de Paiva *et al.* (Eds.)): 124–139. Springer Berlin Heidelberg.

15. Evans, Jonathan St. B. T. (2002). Logic and Human Reasoning: An Assessment of the Deduction Paradigm. *Psychological Bulletin*, Vol. 128, No. 6: 978–996.

16. Geurts, Bart (2003). Reasoning with quantifiers. *Cognition*, 86: 223–251.

17. Geurts, Bart and van der Silk, Frans (2005). Monotonicity and Processing Load. *Journal of Semantics*, 38: 191–258.

18. Geurts, Bart (2007). Existential Import. *Existence: Semantics and Syntax*, I. Comorovski and K. von Heusinger (eds.), 253–271.

19. Geurts, Bart and Katsos, Napoleon and Cummins, Chris and Moons, Jonas and Noordman, Leo (2010). Scalar quantifiers: Logic, acquisition and processing. *Language and Cognitive Processes*, 25(1): 130–148.

20. Goodman, Nelson (1983). Fact, Fiction, and Forecast. *Fourth edition*, Harvard University Press.

21. Grice, P. (1967). Logic and conversation. ch. 2 of *Studies in the Way of Words*, Harvard University Press.

22. Hackl, Martin (2009). On the grammar and processing of proportional quantifiers: *most* versus *more than half*. *Nat Lang Semantics*, 17: 63–98.

23. Hardman, David K. and Payne, Stephen J. (1995). Problem Difficulty and Response Format in Syllogistic Reasoning. *The Quarterly Journal of Experimental Psychology*, 48(A), 4: 945–975.

24. Henle, Mary (1962). On the relation between logic and thinking. *Psychological Review*, Vol. 69, No. 4: 366–378.

25. Icard, Thomas F. and Moss, Lawrence S. (2014). Recent Progress on Monotonicity. *Perspectives on Semantic Representations for Textual Inferences*, CSLI Publications: 167–194.

26. Isaac, Alistair M. C. and Szymanik, Jakub and Verbrugge, Rineke. (2014). Logic and Complexity in Cognitive Science. *Johan van Benthem on Logic and Information Dynamics, Outstanding Contributions to Logic 5*, A. Baltag and S. Smets (eds.), Springer International Publishing Switzerland: 787–824.

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

28. Johnson-Laird, P.N. (1997). Rules and Illusions: A Critical Study of Rips's *The Psychology of Proof*. *Minds & Machines*, 7: 387–407.

29. Johnson-Laird, P.N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences 107.43* (2010): 18243-18250.

30. Katsos, Napoleon and Cummins, Chris and others (2016). Cross-linguistic patterns in the acquisition of quantifiers. *PNAS*, Vol. 113, No. 33: 9244 – 9249.

31. Keenan, Edward L. (2003). Excursions in Natural Logic. Available Online: <http://semanticsarchive.net/Archive/TdiZGUzY/keenan.excursions.pdf>

32. Keenan, Edward L. (2004). Further Excursions in Natural Logic: The Mid-Point Theorems. Available Online: `<http://www.linguistics.ucla.edu/people/`

`keenan/Papers/further\%20excursions\%20in\%20natural\%20logic\%20the\%20mid\%20point\%20problems.pdf>`

**33.** Khemlani, Sangeet and Johnson-Laird, P.N. (2012). Theories of the Syllogism: A Meta-Analysis. *Psychological Bulletin*, 138(3): 427–457.

**34.** Kotek, Hadas and Sudo, Yasutada and Hackl, Martin (2015). Experimental investigations of ambiguity: the case of *most*. *Nat Lang Semantics*, 23: 119–156.

**35.** van Lambalgen, Michiel (1995). Natural Deduction for Generalized Quantifiers. *Quantifiers, Logic, and Language*, Jaap van der Does and Jan van Eijck (eds.): 225–236.

**36.** Mostowski, Marcin (1995). Quantifiers Definable by Second Order Means. *Quantifiers: Logics, Models and Computation, Volume II*, M. Krynicki *et al.* (eds.): 181–214.

**37.** Newstead, S. E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language*, 28: 78–91.

**38.** Newstead, S. E. (2003). Can natural language semantics explain syllogistic reasoning? *Cognition*, 90: 193–199.

**39.** Oaksford, M. and Chater, N. (2001). The probabilistic approach to human reasoning. *TRENDS in Cognitive Science*, Vol. 5, No. 8: 349–357.

**40.** Oaksford, M. and Roberts, L. and Chater, N. (2002). Relative informativeness of quantifiers used in syllogistic reasoning. *Memory & Cognition*,(30)1: 138–149.

**41.** Peirce, Charles Sanders. *The collected papers of Charles Sanders Peirce*, C Hartshorne, P. Weiss, A. Burk (eds.). Cambridge, MA: Harvard University Press.

**42.** Peters, Stanley and Westerståhl, Dag (2006). Quantifiers in Language and Logic. Oxford University Press, Oxford.

**43.** Pietroski, Paul and Lidz, Jeffrey and Hunter, Tim and Halberda, Justin (2009). The Meaning of 'Most': Semantics, Numerosity and Psychology. *Mind & Language*, Vol. 24, No. 5: 554–585.

**44.** Rips, Lance J. (1983). Cognitive Processes in Propositional Reasoning. *Psychological Review*, Vol. 90, No. 1: 38–71.

**45.** Rips, Lance J. (1994). The Psychology of proof: deductive reasoning in human thinking. MIT Press, Cambridge, Massachusetts.

**46.** Roberts, Maxwell J. and Newstead, Stephen E. and Griggs, Richard A. Quantifier interpretation and syllogistic reasoning. *Thinking & Reasoning*, 7(2): 273–204.

**47.** Sánchez Valencia, V. (1991). Studies on natural logic and categorical grammar. Doctoral Dissertation, University of Amsterdam.

**48.** Sells, S.B. (1936). The atmosphere effect: An experimental study of reasoning. *Archives of Psychology*, 29 (Whole No. 200).

**49.** Seuren, Pieter A.M. (2010). *Language from Within Volume II: The Logic of Language.* Oxford University Press.

**50.** Shin, S.-J. (1992). A semantic analysis of inference involving Venn diagrams. *AAAI spring symposium on reasoning with diagrammatic representations*, N.H. Narayanan (ed.): 85–90. Stanford, CA: Stanford University

**51.** Stenning, Keith and van Lambalgen, Michiel (2010). The logical response to a noisy world. *Cognition and Conditionals: Probability and Logic in Human Thinking*, Mike Oaksford and Nick Chater (eds.): 85–101.

**52.** Stenning, Keith and van Lambalgen, Michiel (2012). *Human Reasoning and Cognitive Science*. First Paperback Edition, Cambridge: MIT Press.

**53.** Störring, G. (1908). Experimentelle Untersuchungen über einfache Schlussprozesse. *Archiv für die gesamte Psychologie*, 11: 1–27.

**54.** Strasser, Christian and Antonelli, G. Aldo (2016). Non-monotonic Logic. *The Stanford Encyclopedia of Philosophy (Winter 2016 Edition)*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/logic-nonmonotonic/>

**55.** Strößler, Corina (2017). The Logic of "Most" and "Mostly". *Axiomathes*, DOI 10.1007/s10516-017-9338-2

**56.** Sundholm, Göran (1983). Systems of Deduction. *Handbook of Philosophical Logic Volume I*, Donald Davidson, Gabriël Nuchelmans, Wesley C. Salmon (eds.): 133–188.

**57.** Szymanik, Jakub and Zajenkowski, Marcin (2010). Comprehension of Simple Quantifiers: Empirical Evaluation of a Computational Model. *Cognitive Science*, 34: 521–532.

**58.** Szymanik, Jakub (2016). Quantifiers and Cognition: Logical and Computational Perspectives. *Studies in Linguistics and Philosophy, Volume 96*, Springer International Publishing Switzerland.

**59.** Szymanik, Jakub and Thorne, Camillo (2017). Exploring the relation between semantic complexity and quantifier distribution in large corpora. *Language Sciences*, http://dx.doi.org/10.1016/j.langsc.2017.01.006

**60.** Wason, P. C. (1961). Response to Affirmative and Negative Binary Statements. *British Journal of Psychology*, Vol. 52, No. 2: 133–142.

**61.** Wason, P. C. (1983). Realism and rationality in the selection task. *Thinking and reasoning*, St. B. T. Evans (ed.): 45–75. London: Routledge.

**62.** Westerståhl, Dag (1989). Aristotelian Syllogisms and Generalized Quantifiers. *Studia Logica*, 577–585.

**63.** Wittgenstein, Ludwig (1953/2009). Philosophical Investigations. *Revised 4th edition*, Blackwell Publishing Ltd.

**64.** Zhai, Fangzhou and Szymanik, Jakub and Titov Ivan (2015). Toward Probabilistic Natural Logic for Syllogistic Reasoning. *Proceedings of the 20th Amsterdam Colloquium*, Thomas Brochhagen and Floris Roelofsen and Nadine Theiler (eds.): 468–477.