Journal of Semantics, 36, 2019, 99–163
doi:10.1093/jos/ffy015
Advance Access Publication Date: 09 October 2018
Article

Empty-Set Effects in Quantifier Interpretation

Oliver Bott
Project CiC, SPP XPrag.de, University of Tübingen, 72074 Tübingen, Germany
E-mail: oliver.bott@uni-tuebingen.de

Fabian Schlotterbeck
University of Tübingen, 72074 Tübingen, Germany
E-mail: fabian.schlotterbeck@uni-tuebingen.de

Udo Klein
Empolis Information Management GmbH, Bielefeld, 67657 Kaiserslautern, Germany
E-mail: udomichaelklein@googlemail.com

First version received 28 May 2016; Second version received 23 July 2018; Accepted 29 August 2018

Abstract

This paper proposes a novel, cognitively motivated theory of natural language quantification and presents experimental evidence for it. Taking into account recent findings from number cognition the theory readily explains (1) that the restrictor argument of natural language quantifiers universally has a domain-relativizing function and (2) that monotone decreasing quantifiers are often more difficult to process than increasing quantifiers. In the proposed theory, quantifiers that do not have the empty set as a witness set are simpler to evaluate than those that do (what we call *empty-set quantifier*s). This explains enhanced difficulty of monotone decreasing quantifiers because they are always *empty-set quantifiers*. Furthermore, quantificational complexity is predicted to be extraordinarily high if such an *empty-set quantifier* has to be evaluated in a situation in which its scope is empty (what we call *empty-set situations*). These predictions were tested in three experiments investigating the online comprehension and the verification of quantified sentences. The first experiment established *empty-set effects*, i.e., enhanced difficulty during the online interpretation and the evaluation of empty-set quantifiers relative to non-empty-set quantifiers, particularly during the evaluation of empty-set situations. The second experiment supports our claim that empty-set effects have to be distinguished from monotonicity effects. Empty-set effects were observed for the non-monotone Boolean combination of quantifiers *none or three dots* relative to *one or three dots*. The third experiment shows that the proposed theory of quantificational complexity is not limited to simply quantified sentences but can account for doubly quantified sentences, too. It was manipulated whether the first and/or the second quantifier were empty-set quantifiers. The experiment shows that the difficulty of empty-set quantifiers adds up in a cumulative fashion – a finding only expected under a 'recursive' version of the proposed theory.

1 INTRODUCTION

Quantifiers have been of core interest to philosophers, linguists and logicians (for a comprehensive review, see, e.g., Peters & Westerståhl 2006), and psychologists have collected a wealth of data on how quantified statements are interpreted and what makes them difficult to comprehend (see, e.g., Just & Carpenter 1971; Johnson-Laird 1983; Johnson-Laird & Bara 1984; Sanford *et al.* 1994; Tunstall 1998; Paterson *et al.* 2009; Urbach & Kutas 2010; Nieuwland 2016, among others). We may thus expect to find formal theories of quantifier interpretation able to account for processing data obtained in psychological studies on quantification. Formal semantic theories on quantification have, however, been developed without much consideration of how quantifiers are actually processed, even though semanticists have started to propose semantic theories striving for cognitive reality (e.g., Szymanik 2016). The present paper proposes a novel and cognitively motivated formal theory of quantifier interpretation which is able to comprehensively account for a number of experimental findings from the processing literature. The proposed theory provides us with a clear notion of quantificational complexity. Furthermore, its cognitive plausibility is independently motivated by recent research from number cognition.

This paper develops an algorithmic theory of quantifier interpretation roughly in the sense of Moschovakis (1994): The *sense* of an expression is the algorithm which computes its *denotation*. Meaning understood in this way is inseparably linked to verification procedures. Since only a limited number of the theoretically possible verification procedures may be cognitively plausible this view forces us to formulate semantic theories in such a way that cognitive constraints are directly reflected in the semantics. In the proposed theory quantificational complexity is modeled at two levels. Firstly, *the algorithms* corresponding to two quantified statements can differ in complexity, that is, one algorithm can be inherently more complex than the other. We take this kind of complexity to affect the difficulty of computing a semantic representation during the comprehension of quantified statements presented out of context. Secondly, depending on the context against which a quantified sentence is evaluated, *the execution* of a given algorithm can vary with respect to the number and kind of steps required in order to compute its truth value. This corresponds to the complexity of a given instance of the verification problem. The experiments reported below will specifically target these two aspects by looking separately at the comprehension of quantified statements and their evaluation in a concrete situation.

A classic example of an algorithmic theory of natural language quantification are the semantic automata proposed by van Benthem (1986).[1] Automata theory has inspired a growing body of psycho- and neurolinguistic work on quantifier processing (see, e.g., McMillan *et al.* 2005; Szymanik 2016, and the review in Section 3.1). What sets our proposal apart from the automata model is the assumption of a general asymmetry in processing *positive* vs. *negative instances of a predicate*, i.e. entities that are or are not

1 Although the origins of the algorithmic approach date back further at least to Church (1936)'s seminal paper on the λ-calculus. We would like to thank an anonymous reviewer for pointing this out.

in the denotation of a predicate. This asymmetry readily allows us to account for effects of monotonicity – or rather empty-set effects as we argue – not readily accounted for by the automata model.

The assumed asymmetry between positive and negative predicate instances naturally connects our proposal with yet another area of work on quantifier processing probing into the relation between semantic representations of quantifiers and their processing as realized in the cognitive architecture (cf. Hackl 2009; Pietroski *et al.* 2009; Lidz *et al.* 2011; Tomaszewicz 2011; Kotek *et al.* 2015; Hunter *et al.* 2016). Lidz *et al.* (2011) proposed the Interface Transparency Thesis (ITT) stating that "verification procedures employed in understanding a declarative sentence are biased towards algorithms that directly compute the relations and operations expressed by the semantic representation of that sentence." Since cognitive limitations may plausibly constrain the verification procedures viable in semantic interpretation, the ITT lets us expect that specifications of truth conditions for declarative sentences are constrained accordingly.

In this paper we will focus on one specific cognitive operation that can be expected to be exceptionally costly, namely the evaluation of the empty set as having cardinality zero. Research on number cognition suggests that this operation is qualitatively different from the evaluation of positive numerosities (see Section 1.1). We therefore expect that the verification of quantifiers is biased towards procedures that do not depend on this operation. This suggests that quantifiers that can be verified with an algorithm that depends solely on positive numerosity should be less difficult to process than quantifiers which require the evaluation of the empty set. Henceforth, we will refer to the evaluation of the empty set as *empty-set situations* and to processing difficulty related to their evaluation as *empty-set effects*. Quantifiers that are true in such a situation will be referred to as *empty-set quantifiers*, and quantifiers that are false will be referred to as *non-empty-set quantifiers* (cf. Definition 9.9).

The paper is structured as follows. The next subsection provides an empirical motivation for the just mentioned asymmetry between the evaluation of the empty set and sets with positive cardinality. Section 2 informally summarizes our theoretical proposal, which makes crucial use of this distinction. Section 3 compares the proposed theory to alternative algorithmic theories of natural language quantification and relates it to existing experimental studies on quantifier interpretation. Section 4 summarizes the central hypotheses tested in the experiments and derives predictions about empty-set effects. Sections 5–7 present three experiments testing these predictions. Section 8 contains a general discussion of the empirical findings. A formal version of the theory is worked out in Section 9. The final section concludes.

1.1 *Cognitive motivation from number cognition*

A substantial body of work on number cognition has investigated the processing of number and numerosity, another term for the cardinality of a set, and the underlying mental concepts and representations (see, e.g., Dehaene 1997; Feigenson *et al.* 2004; Dehaene & Nieder 2009). Within this body of work a growing number of studies focussed on the evaluation of the empty set and the according concept of the number zero. This research points at a qualitative difference between processing numerosity zero and numerosities corresponding to the positive integers (see the review in Ramirez-Cardenas *et al.* 2016). We take this difference as a crucial, cognitive motivation for the quantification theory presented in Sections 2 and 9 (cf. Schlotterbeck 2017, especially pp. 184–192).

The human system of number and quantification is deeply rooted in a biologically primitive, non-verbal system of number cognition that can be found in non-human animals as well as human infants and adults regardless of their cultural background (Dehaene & Nieder 2009). This system computes and represents quantity information about collections of objects irrespective of other properties they have. It approximates the number of objects present in sensory input.

Since the seminal study of Moyer & Landauer (1967) it is assumed that the mental representations of numerical information are noisy and use an analog format. The main motivation for this assumption are the well-known distance and size effects on proportions of errors and RTs in number comparison or discrimination tasks. The distance effect corresponds to larger proportions of errors and longer RTs when two numbers are closer than when they are more distant from each other. The size effect corresponds to larger proportions of errors and longer RTs for larger numbers relative to smaller numbers when their numerical distance is held constant. These well-established effects are in turn commonly used as a tool for studying the cognitive processes underlying number cognition, which have been described down to the level of single neurons accounting for the firing rates and tuning curves of numerosity encoding neurons in the primate brain (Nieder *et al.* 2002). Size and distance effects are not only observed in behavioral measures but can be detected in the firing rates of these neurons, too. The existing studies provide strong evidence for a neurally implemented logarithmically compressed mental number line with number receptors working in an analog and noisy fashion (Nieder 2016b). This research shows that the brain contains neurons dedicated to detecting specific numerosities and these 'number receptors' receive direct activation from sensory input while abstracting away from other properties such as form, size, colour, and so forth. Computational models of number cognition incorporate these numerosity encoding neurons in their architecture. For instance, Verguts & Fias (2004) presented a connectionist neural model of basic numerical abilities extending earlier work by Dehaene & Changeux (1993). Their model includes number detectors that are activated by sensory input either in the form of a layout of objects or in the form of number symbols that have been associated with numerosity during learning. Acquisition of numerical skills is thought to start with learning to count small numbers of items based on abstraction over sensory input and then progressively use this counting procedure to comprehend arbitrary numbers in symbolic form Wiese (2003).

The number zero differs in a crucial respect from the just described processing of positive numbers: It encodes the absence of objects rather than their presence and is thus detached from sensory experience. Recent studies show that the concept of numerosity zero is in fact processed quite differently from positive numerosities (Nieder 2016a). Although the number concept zero is eventually incorporated into the number line – as evidenced by distance effects – it proves not only to be special with respect to its late invention in human history, but also its late emergence in human development, its limited evolution in animals, and its special representation in the brain. To just mention one study, Ramirez-Cardenas *et al.* (2016) report an experiment using a discrimination task in rhesus monkeys for displays including zero to four objects while simultaneously recording activity from single neurons in the primates' parietal and frontal lobes. For both recording sites the authors found neurons tuned to positive numerosities as well as empty sets. However, while distance effects for positive integers were observed in the firing rates of neurons for both recording sites, 'empty-set neurons' gave rise to reliable distance effects only in the prefrontal cortex. The authors interpret this finding as suggesting that the representation of numerosity zero results from a fundamentally different hierarchically organized processing routine for zero than for positive integers. Whereas positive integers are externally triggered by sensory information with number detectors in modality specific sensory areas, the representation of numerosity zero is generated internally in hierarchically ‘higher’ parts of the cognitive system. The generation process can be conceived of as an inference from nothing to something: Neurons tuned to detecting the absence of sensory input give rise to the activation of number detectors tuned to numerosity zero Nieder (2016a).

The outlined difference between positive cardinalities and zero can be expected to have consequences for processing natural language quantifiers such as *at most three*. In contrast to quantifiers such as *between one and three* which can be represented by a mere combination of ‘number detectors’, that is detectors for *one*, *two*, and *three*, the empty-set quantifier *at most three* does not allow for such a simple representation. Instead, it involves checking whether a given set consists of either between one and three objects or zero elements. Given the just mentioned functional differences between the evaluation of positive numerosities and numerosity zero, the quantifier *at most three* can thus be expected to simultaneously involve two checking procedures of rather different kinds. The proposed quantification theory is built on this distinction. It crucially relies on the difference between having to evaluate the empty set as opposed to non-empty sets.

2 A NOVEL THEORY OF QUANTIFICATION – INFORMALLY

In line with the just mentioned work from number cognition we propose that the evaluation of an empty-set situation is different from the evaluation of positive numerosities because it involves an inference from the absence of predicate instances to the truth of a quantified statement. However, whether this kind of inference is actually required for the truth evaluation of quantified statements crucially depends on the semantic properties of the involved quantifiers. We propose that natural language quantificational determiners fall into two classes. While the first class can be evaluated without such an inference, the second class crucially involves the evaluation of empty-set situations.

Here is an informal summary of the basic ideas worked out formally in Section 9. In contrast to the treatment of quantificational determiners as binary relations over sets (called the restrictor and the scope set) in Generalized Quantifier Theory (Barwise & Cooper 1981; van Benthem 1986; Keenan & Stavi 1986), the present theory identifies them with unary functions that map a restrictor set to its set of witness sets– think of witness sets as those subsets of the restriction to which the restricted quantifier applies truthfully (a formal definition of these functions, which we call *w-functions*, is given in Section 9, Definition 9.5). For example, the determiner *exactly three* in *exactly three boys* maps the set of boys to those subsets with cardinality three. To give another example, the determiner *no* in *no woman* maps the set of women to a single set, the empty set. One – in our view very welcome – consequence of this treatment of determiners is that their interpretation only depends on the restrictor set and the subsets that are among their witnesses. Anything else can be safely ignored. This way, semantic universals such as the domain-relativizing function of the restriction are hard-wired into the system without the need to stipulate additional constraints such as conservativity (Conserv), extension (Ext), and isomorphism closure (isom) (Barwise & Cooper 1981; Keenan & Stavi 1986; von Fintel & Matthewson 2008); cf. Definitions 9.1 and 9.2 of Section 9.

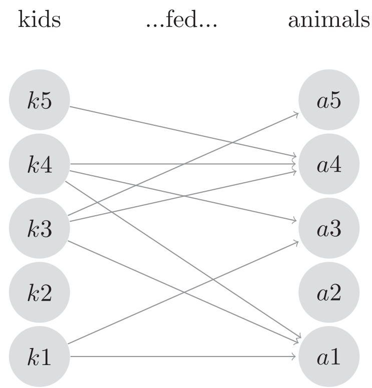


Figure 1 Sample situation for the sentences discussed in Section 2.

To see how this conception of quantification can be applied to verification procedures, consider the situation depicted in Figure 1. The 'sensory input' can be summarized as follows. There is a set, $\{k1, \ldots, k5\}$, consisting of five kids and another set, $\{a1, \ldots, a5\}$, consisting of five animals. These two sets are related by a binary relation, $\{\langle k1, a1\rangle, \langle k1, a3\rangle, \langle k3, a1\rangle, \langle k3, a4\rangle, \langle k3, a5\rangle, \langle k4, a1\rangle, \langle k4, a3\rangle, \langle k4, a4\rangle, \langle k5, a4\rangle\}$ encoding which kid fed which animal. This is all the 'positive information' available in the depicted situation. Note that we have not stated explicitly which individuals do NOT participate in the feeding relation, e.g. that $k2$ fed no animal.

Based on this representation, we can start to evaluate quantified statements, e.g., the doubly quantified sentences in (1-a) and (1-b).

(1) a. Exactly three kids fed more than one animal.
b. Most kids fed more than one animal.

Both of these sentences are true in the depicted situation. Here is a sketch of how truth evaluation could proceed in a rather simple manner. Let us for now assume that both sentences are interpreted with surface scope, that is the linearly first quantifier takes scope over the second.[2] In order to evaluate quantified sentences we introduce an operation henceforth referred to as **simple expansion** (**s-exp**, see Definition 9.18 for a formal definition). The general form of this operation involving a single rule is given in (2). In order to keep the description concise, the notation $\sigma^{[n/x]}$ is used, which simply denotes the tuple σ' that is identical to σ except that the n-th element is replaced by x.

(2) **Simple expansion** (**s-exp**) **rule** for a quantifier Q_i as the ith argument of an n-ary predicate P (with $1 \leq i \leq n$): For each tuple $\sigma \in P$ collect the set $\{x : \sigma^{[i/x]} \in P\}$. Then, intersect this set with the restriction of Q_i and check whether the result is a *witness set* of Q_i. If so, add $\sigma^{[i/Q_i]}$ to the predicate denotation of P.

Let us look at the examples in (1-a) and (1-b) to illustrate how **s-exp** works. Quantifiers are added to the verb denotation successively starting with the narrow-scope quantifier *more*

2 Below we will come back to the relative scope of quantifiers (cf. Definition 9.25ff. in Section 9).

than one animal. The denotation of *fed* from above can be straightforwardly expanded applying the just outlined operation. However, the operation only makes sense once the concept of an n-ary predicate has slightly been adjusted: instead of signifying a relation over individuals, predicates must be defined in such a way that they can also relate quantifiers to individuals or to each other (for a related proposal see Hendriks 1993, and compare the definition of an *n-ary predicate* in Definition 9.10). Once such a notion of an n-ary predicate is in place, we proceed as follows:

For each kid k_i we check whether k_i fed more than one animal, i.e. whether the denotation of *fed* contains at least two pairs with kid k_i as its first element. Whenever this is true, the pair $\langle k_i,$ *more than one animal*$\rangle$ is added to the denotation of *fed*. Iterating through all of the kids, the **s-exp** operation results in adding three pairs to the denotation of *fed*: $\langle k1,$ *more than one animal*$\rangle$, $\langle k3,$ *more than one animal*$\rangle$, and $\langle k4,$ *more than one animal*$\rangle$. Note that this information was already implicitly contained in the denotation of the predicate but it now is made explicit which of the kids fed more than one animal.

The next step is to further expand the predicate denotation with the subject quantifier. First consider the non-monotone quantifier *exactly three kids* in (1-a). Now check whether the set $\{k1, k3, k4\}$ (consisting of those individuals x for which $\langle x,$ *more than one animal*$\rangle$ has been added to *fed*) is among the set of witness sets of *exactly three kids*. Since the cardinality of this set is three, and all subsets of *kids* with cardinality three are among the set of witness sets of *exactly three kids*, simple expansion turns out to be successful again, and the pair $\langle$*exactly three kids*, *more than one animal*$\rangle$ is also added to the denotation of *fed*. This step can be identified with verification (cf. Definition 9.27): A doubly quantified sentence with a subject quantifier Q_1 and an object quantifier Q_2 is true just in case the sequence of expansion operations as determined by the scope reading of the sentence results in adding the pair $\langle Q_1, Q_2\rangle$ to the predicate denotation.

Example (1-b) can be dealt with similarly. It has to be checked whether $\{k1, k3, k4\}$ is among the witness sets of *most kids*. Since in the above scenario the set of witness sets of *most kids* consists of all the subsets of the kids that have a greater cardinality than two, the pair $\langle$*most kids*, *more than one animal*$\rangle$ can be added to the denotation of *fed* and the sentence is thus evaluated *true*.[3]

Nevertheless we would like to point out that the checking procedure is slightly more involved than in the case of *exactly three kids*. For the latter only the intersection of the restrictor and the scope set is relevant, whereas proportional quantifiers such as *most kids* also require the evaluation of the cardinality of the restrictor set as a whole.

3 The just stated verification procedure is in line with the proposed representation of the quantifier *most* advocated by Lidz *et al.* (2011). It implements a strategy based on cardinality subtraction and comparison of just two sets: The set of kids was part of the representation we started with and the interpreter can therefore directly access its numerosity. The only other numerosity required is that of the set $\{k1, k3, k4\}$ constructed when expanding the object quantifier.
Furthermore, note that the truth-conditionally equivalent *one-to-one-plus* strategy discussed and dismissed by Pietroski *et al.* (2009) is not consistent with the evaluation of the example via **s-exp**. This is because the *one-to-one-plus* strategy is based on establishing a one-to-one correspondence between those kids that fed more than one animal and those that did not. It thus crucially presupposes the explicit representation of those kids that did not feed more than one animal. Obviously, however, this set is by assumption not part of the assumed situation model.

Having illustrated the simple expansion operation for the linear scope readings of the sentences we would like to briefly flesh out how inverse scope is implemented in our theory (cf. Definitions 9.25–9.27, and the discussion of Example (27) in Section 9). To evaluate the inverse scope reading, it suffices to reverse the order of expansion steps: Instead of starting with expansion of the object quantifier, the subject quantifier has to be expanded first, or, to put it differently, we have to start with the quantifier with narrowest scope. This shows that it is not sufficient to consider whether a given predicate can be expanded with a tuple involving the relevant quantifiers, but that it is crucial to consider the order of expansion steps (i.e. the scope reading) with which this expansion is achieved. The outlined procedure is applicable to an arbitrary number of quantifiers under any possible scope reading.

The outlined account is explicitly restricted to scope readings of multiply quantified sentences which contain 'simple' (type $\langle 1,1\rangle$, monadic) quantifiers or Boolean combinations thereof, which are also of type $\langle 1,1\rangle$ (cf. Definition 9.5 and the following paragraphs). However, other readings of multiply quantified sentences exist that are not covered by the present framework. In Generalized Quantifier Theory, the iteration of quantifiers used to account for their scope readings is only one form of higher order (polyadic) quantification (see, e.g., Szymanik 2010). Other forms include cumulative or collective interpretations such as (3-a), resumption illustrated in (3-b), the branching quantification reading of (3-c), and the various interpretations of reciprocal sentences with quantifier antecedents such as (3-d). The first three examples are adopted from Szymanik (2010).

(3) a. Eighty professors taught sixty courses at ESSLLI.
b. People usually are grateful to firemen who rescue them.
c. Most villagers and most townsmen hate each other.
d. At least four dots are connected with each other.

While we have nothing to say about resumption, branching quantification, and the interpretations of reciprocals, we would like to briefly comment on collective quantification in (3-a). This example is most naturally understood as neither saying that there were eighty times sixty ESSLLI courses (the linear scope interpretation), nor that there were sixty courses with eighty instructors each (the inverse interpretation), but that eighty professors taught a total of sixty ESSLLI courses and that each course had at least one instructor. Szymanik (2010, p. 226-227, Definition 18) proposed an analysis of cumulation (*Cum*) within Generalized Quantifier Theory formalizing the just given paraphrase (but see Robaldo *et al.* 2014, for a different approach based on witness sets, and for more discussion on collective interpretations). This analysis, which can also be found in Peters & Westerståhl (2006, p. 351), shows that cumulation can be defined in terms of iteration (*It*); A and B correspond to the restrictor sets of the quantifiers Q and Q', and R is a transitive predicate with the inverse relation R^{-1}, e.g., *were taught by*:

(4) $Cum(Q,Q')[A,B,R] \Leftrightarrow It(Q,Some)[A,B,R] \wedge It(Q',Some)[B,A,R^{-1}]$

Obviously, the suggested operation on the right-hand side in (4) can be mimicked in the present account using the simple expansion operation and applying it to (3-a). The procedure is straightforward. First, we loop through the set of professors, i.e. the restrictor set of *eighty* and check which of them stand in the relation denoted by *taught* to some ESSLLI course, i.e. which professors gave a course at ESSLI. The resulting set is then checked whether it is among *eighty*'s witness sets. If this is the case, we can move on to the second conjunct of (4) and check which of the courses were actually taught by some professor. If the resulting set is among the witness sets of *sixty*, simple expansion evaluates the second conjunct of (4) to be true and propositional logic yields that the cumulative reading is true. Otherwise it is evaluated as false. Whether such a procedure is cognitively realistic goes well beyond the scope of the present proposal and has to be left to future research. A point worth mentioning, though, is that the distinction between simple expansion and complex expansion to be defined in the next section, should also be relevant for cumulative construals if we think of collective readings in the way as suggested by (4). Depending on whether the sentence contains empty-set quantifiers or not, an extension of the proposed account in this direction would predict empty-set effects in collective readings, too.

2.1 *Limits of the simple expansion operation*

Crucially, in all the cases discussed so far we only made use of positive information, i.e. objects or relations present in the ‘sensory input’. For instance, in the first derivation we only expanded the predicate *fed* with pairs consisting of a kid and the quantifier *more than one animal* if this pair constituted a positive instance of this relation. All the negative instances such as the kids $k2$ and $k5$ in Figure 1 were, however, systematically ignored in the evaluation process. This raises the question under which circumstances, or rather for which determiners, the outlined procedure yields the correct results, and under which circumstances it fails. A central proposition (Proposition 9.28) of the present theory is that the **s-exp** algorithm works for all possible scope readings of multiply quantified sentences with an arbitrary number of quantifiers as long as none of these quantifiers has the empty set among its witness sets. Examples of empty-set quantifiers (cf. Definition 9.9) are of course *no dog*, but also monotone decreasing quantifiers such as *fewer than three dogs*, and non-monotone quantifiers resulting from Boolean combinations such as *none or three dogs*. In fact, reconsidering examples (1-a) and (1-b) once more shows that all quantifiers discussed so far were non-empty-set quantifiers; none of them had the empty set among their witness sets.

The proposed **s-exp** algorithm does not work in the case of empty-set quantifiers. This becomes evident in the following example, again evaluated in the situation shown in Figure 1.

(5) Exactly three kids fed less than three animals.

This sentence is true under a linear scope construal because exactly three kids, namely $k1$, $k2$, and $k5$, fed fewer than three animals. However, the **s-exp** algorithm evaluates it as false. Simple expansion of the object quantifier *less than three animals* has only access to the kids $k1$, $k3$, $k4$, and $k5$ since only these kids are explicitly represented in the denotation of *fed*. So this is what we get: For $k1$ the set $\{a1, a3\}$ is evaluated, which is a witness set of *less than three animals* and therefore **s-exp** adds the pair $\langle k1$, *less than three animals*$\rangle$ to the denotation of *fed*. Similarly, for the pair $\langle k5$, *less than three animals*$\rangle$. The predicate denotation is not expanded with the pairs $\langle k3$, *less than three animals*$\rangle$ or $\langle k4$, *less than three animals*$\rangle$ because the sets $\{a1, a4, a5\}$ and $\{a1, a3, a4\}$ are not witness sets of *less than three animals*. Thus, expansion of the subject quantifier *exactly three kids* ends up with the set $\{k1, k5\}$. Consequently, the predicate is not expanded with the pair $\langle$*exactly three kids*, *less than three animals*$\rangle$, and the sentence is incorrectly evaluated as false.

The problem and its solution are obvious. For an empty-set quantifier, such as *less than three animals*, we have to keep track of individuals, such as $k2$, NOT feeding any animals. To do so, we first enrich the representation of the predicate *fed* by adding the pairs that are not among the positive instances (we call this kind of enriched representation *polarity relation*, cf. Definition 9.13). These negative instances are marked with a minus sign, whereas positive instances are henceforth marked with a plus sign. The resulting binary relation *fed* corresponds to the following set:

(6) $\{\langle k1, a1\rangle^+, \langle k1, a2\rangle^-, \langle k1, a3\rangle^+, \langle k1, a4\rangle^-, \langle k1, a5\rangle^-,$
$\langle k2, a1\rangle^-, \langle k2, a2\rangle^-, \langle k2, a3\rangle^-, \langle k2, a4\rangle^-, \langle k2, a5\rangle^-,$
$\langle k3, a1\rangle^+, \langle k3, a2\rangle^-, \langle k3, a3\rangle^-, \langle k3, a4\rangle^+, \langle k3, a5\rangle^+,$
$\langle k4, a1\rangle^+, \langle k4, a2\rangle^-, \langle k4, a3\rangle^+, \langle k4, a4\rangle^+, \langle k4, a5\rangle^-,$
$\langle k5, a1\rangle^-, \langle k5, a2\rangle^-, \langle k5, a3\rangle^- \langle k5, a4\rangle^+, \langle k5, a5\rangle^-\}$

Having labeled all the tuples as positive or negative predicate instances, respectively, we can go through all the elements of the restrictor sets of the quantifiers in the sentence. Instead of the simple expansion operation we now introduce what we refer to as the **complex expansion operation** (**c-exp**, formally defined in Definition 9.21):

(7) **Complex expansion (c-exp) operation** for a quantifier Q_i as the ith argument of an n-ary predicate P (with $1 \leq i \leq n$): For each tuple σ^+ or $\sigma^- \in P$ collect the set $\{x : \sigma^{+[i/x]} \in P\}$, respectively $\{x : \sigma^{-[i/x]} \in P\}$. Then, intersect this set with the restriction of Q_i and execute the following instructions:

1. If the result was derived from a positive tuple and is a witness set of Q_i, add $\sigma^{+[i/Q_i]}$ to the predicate denotation.
2. If the result was derived from a positive tuple and is not a witness set of Q_i, add $\sigma^{-[i/Q_i]}$ to the predicate denotation.
3. If the result is the entire restriction and was derived from a negative tuple and Q_i has the empty set among its witness sets, add $\sigma^{+[i/Q_i]}$ to the predicate denotation.
4. If the result is the entire restriction and was derived from a negative tuple and Q_i does not have the empty set among its witness sets, add $\sigma^{-[i/Q_i]}$ to the predicate denotation.

Comparing the **c-exp** operation in (7) to the **s-exp** operation in (2) shows that the latter is actually a proper part of the former. Complex expansion just introduces additional clauses that explicitly deal with empty-set situations. To illustrate how complex expansion works, let us go through the example in (5) given the situation shown in Figure 1 again.

Starting with the quantifier with narrowest scope, it is checked for each kid whether it fed *less than three animals*. Kid $k1$ fed animals $a1$ and $a3$. Since the set $\{a1, a3\}$ consists of only two animals, it is a witness set of *less than three animals* and clause 1 applies. Therefore, the pair $\langle k1,$ *less than three animals*$\rangle^+$ is added to the predicate denotation. Kid $k2$ did not feed any animals, and, since *less than three animals* is an empty-set quantifier, clause 3 applies. As a result, the pair $\langle k2,$ *less than three animals*$\rangle^+$ is added to the predicate denotation. Kid $k3$ fed animals $a1$, $a4$, and $a5$. Because the corresponding set is not among the witness sets of *less than three animals* clause 2 applies, and the pair $\langle k3,$ *less than three animals*$\rangle^-$ is added to the predicate denotation. Similarly for $\langle k4,$ *less than three animals*$\rangle^-$.

For the last kid $k5$ expansion is achieved via clause 1, and the result is $\langle k5$, *less than three animals*$\rangle^+$. The evaluation of the narrow scope object quantifier can thus be summarized as follows:

(8) $\{\langle k1,$ *less than three animals*$\rangle^+$, $\langle k2,$ *less than three animals*$\rangle^+$, $\langle k3,$ *less than three animals*$\rangle^-$, $\langle k4,$ *less than three animals*$\rangle^-$, $\langle k5,$ *less than three animals*$\rangle^+\}$

Expansion with the subject quantifier proceeds exactly as in example (1-a) above. The pair $\langle$*exactly three kids*, *less than three animals*$\rangle^+$ is added to the predicate denotation via clause 1. In other words, the linear scope reading of (5) is evaluated as true. Finally, let us briefly consider the following sentence.

(9) No kid fed more than three animals.

Here the evaluation of the object quantifier – again under its narrow scope reading – proceeds via clause 2 of the **c-exp** operation for kids $k1$, $k3$, $k4$, and $k5$, and with clause 4 for kid $k2$ because *more than three kids* is a non-empty-set quantifier. Hence, all pairs involving individual boys as their first argument and the object quantifier *more than three animals* are marked as negative predicate instances. Accordingly, expansion with the wide-scope empty-set subject quantifier *no kid* will proceed via clause 3, and sentence (9) is correctly evaluated as true.

2.2 *From expansion operations to quantificational complexity*

The just sketched theory of quantifier interpretation has clear processing consequences. The interpretation and evaluation of quantified statements only involving non-empty set quantifiers should be less complex than that of quantified statements involving empty-set quantifiers. This is for two reasons. Firstly, in contrast to the **s-exp** operation, the **c-exp** operation requires to keep track not only of the positive instances of a predicate but also of its negative instances. And secondly, clause 3 embodies an inference from 'negative information' to 'positive information', similarly to what has been suggested in the above-mentioned studies on number cognition. With reference to the findings of these studies, our proposal leads us to expect that the evaluation of empty-set quantifiers in empty-set situations – situations that require verification via clause 3 – is a major source of difficulty in the processing of quantifiers.

The proposed theory immediately explains one basic finding in work on quantifier processing. It is commonly observed that (right) monotone decreasing, or (DE), quantifiers such as *fewer than three kids* are generally harder to process than (right) monotone increasing, or (UE), quantifiers such as *more than three kids* (see Definition 9.8 for formal definitions of monotonicity properties). One reason for enhanced processing difficulty of monotone decreasing quantifiers is that they are always empty-set quantifiers. By contrast, typical instances of monotone increasing quantifiers are non-empty-set quantifiers. As a consequence, monotone decreasing quantifiers cannot be evaluated using the **s-exp** operation but call for the **c-exp** operation while monotone increasing quantifiers can be evaluated based on simple expansion. However, under our account monotonicity is merely correlated with quantificational complexity. This is because there are monotone increasing quantifiers that are empty-set quantifiers. Consider, for instance *all but at most twenty* *students*, which is monotone increasing but also an empty set quantifier.[4] Similarly, we also predict that monotone decreasing quantifiers are generally harder than non-monotone quantifiers such as *exactly three kids*. However, there are also non-monotone empty-set quantifiers, for example Boolean combinations like *none or exactly three kids*.

So far, we have only distinguished between two broad classes of quantificational statements. The first class consists of statements with an iteration of non-empty-set quantifiers, which can be verified with the **s-exp** operation. The second class consists of iterations of quantifiers including at least one empty-set quantifier, which cannot be evaluated by simple expansion, but must be verified using the **c-exp** algorithm. However, the general definition of the expansion operation allows us to consider alternative scenarios, too. It is, for instance, possible to define a sequence of expansion operations for a multiply quantified statement in such a way that for each quantifier the interpretation system chooses the simplest expansion operation, i.e. decides between **s-exp** and **c-exp** in a quantifier-by-quantifier fashion. In order to verify the surface scope reading of sentence (9) given above, the interpretation system could, for example, choose to use a sequence of an **s-exp** operation followed by an **c-exp** operation. In the first step, the object quantifier *more than three animals* would be evaluated using the **s-exp** operation to expand the second argument position of the verbal predicate. Afterwards, in the second step, the **c-exp** operation would have to be used to evaluate the subject quantifier *no kid* by expanding the first argument position of the predicate. In terms of complexity, this sequence of two expansion operations lies between an algorithm that relies on two applications of the **s-exp** operation and an algorithm that uses two applications of the **c-exp** operation. Note that in the case of multiply quantified sentences with at least one empty-set quantifier such a 'mixed procedure' still requires the full encoding of the positive and the negative instances of a predicate as well as the systematic expansion of all restrictor elements with positively and negatively labeled tuples, respectively. On top of this added complexity the expansion of a non-empty-set quantifier in such an iteration sequence should be easier than that of an empty-set quantifier because in contrast to the latter, safe evaluation of non-empty-set quantifiers can be achieved on the sole base of clause 1 and *negation as failure* (as known from logic programming, see e.g. Lloyd 1987) in those cases in which it is not applicable.

Below, we will contrast a 'globalist version' with the just outlined 'recursive version' of our theory (cf. Hypotheses 4.2 and 4.3). The difference between these two views can be summarized as follows, again assuming linear scope with quantifier Q_1 scoping over Q_2 and *es* short for *empty set*.

(10)

	globalist version	recursive version
non-es $Q1$ / *non-es* $Q2$	**s-exp**	**s-exp**($Q2$) ⇒ **s-exp**($Q1$)
es $Q1$/*non-es* $Q2$	**c-exp**	**s-exp**($Q2$) ⇒ **c-exp**($Q1$)
non-es $Q1$ / *es* $Q2$	**c-exp**	**c-exp**($Q2$) ⇒ **s-exp**($Q1$)
es $Q1$ / *es* $Q2$	**c-exp**	**c-exp**($Q2$) ⇒ **c-exp**($Q1$)

4 The sentence *all but at most twenty students knew the answer* would, strictly speaking, be a truthful description of a situation in which none of a class of fifteen students knew the answer to the relevant question. Of course, this sentence would at the same time be a completely uninformative statement about this particular class and would thus be unlikely to be used to describe this situation. But this is for pragmatic reasons.

3 ALTERNATIVE THEORIES AND FINDINGS FROM EXPERIMENTAL STUDIES ON QUANTIFIER INTERPRETATION

Having introduced the theory we first compare it to other theories of quantifier interpretation. Afterwards, we will review empirical findings from the psycholinguistic literature on quantifier processing and show that the sketched theory can accommodate a large number of them.

3.1 *Alternative theories*

Other theories of quantifier interpretation do not readily capture the complexity differences between DE and UE quantifiers (an exception is the theoretical proposal developed by Schlotterbeck (2017, ch. 4–7), who also compares various prominent theoretical alternatives). One famous example of an algorithmic model of quantification is the automata model of van Benthem (1986) and further developed, for example, by Mostowski (1998), Szymanik (2009), Steinert-Threlkeld & Icard (2013) and Steinert-Threlkeld (2016). In a nutshell, these authors conceive of quantifiers as *decision problems* in the automata theoretic sense Hopcroft & Ullman (1979) and study them from a computational perspective. A number of theorems characterize the computational resources that are needed for the truth evaluation of quantified sentences. The most famous example, due to van Benthem (1986), identifies first-order definable quantifiers with acyclic deterministic finite state automata. From this it follows that the quantifier *most*, which cannot be defined in first-order logic, cannot be evaluated by such a device. Instead a push-down automaton is required, i.e. an automaton equipped with a rudimentary memory device (see also our discussion of *most* vs. *exactly three* in Section 2). Most of this work is limited to statements that contain one quantificational determiner. Recently, however, Steinert-Threlkeld & Icard (2013) extended the automata model from simple quantificational determiners to their iteration – the type of truth conditions discussed in the previous section.

Several authors hypothesized that theoretical distinctions like these affect how quantified sentences are processed. This hypothesis gained support from a number of experimental studies (e.g. McMillan *et al.* 2005, 2006; Szymanik 2009; Szymanik & Zajenkowski 2010, 2011; Zajenkowski *et al.* 2011; Zajenkowski & Szymanik 2013; Zajenkowski *et al.* 2014).

The automata model does, however, not account for one crucial aspect of the processing of quantifiers, namely that DE quantifiers are inherently more difficult to process than UE quantifiers (see Subsection 3.2.3 for a summary of empirical findings supporting this claim). In the automata model, DE and UE quantifiers are treated in a strictly parallel fashion, both in terms of the kind of automata (e.g. deterministic finite state vs. pushdown automaton) and in terms of the number of states and transitions involved in minimal pairs such as *more than n* and *fewer than n-1*. This parallelism is illustrated in Figure 2 showing the simplest automata for *more than three* and *fewer than four*.

To explain the enhanced difficulty of upward vs. downward monotone modified numerals, Szymanik (2016) suggested that "'passing through accepting states' is more difficult than 'passing through rejecting states"' (p. 73). This would predict that montonicity effects should increase with the numeral because larger numerals require more states. This prediction seems implausible to us, but it remains an empirically open question. Comparing the Boolean combination *none or three*, an empty-set quantifier, with the non-empty-set quantifier *one or three* in Panels (c) and (d) of Figure 2 it becomes evident that these quantifiers correspond to strikingly similar automata: They have the same number of states

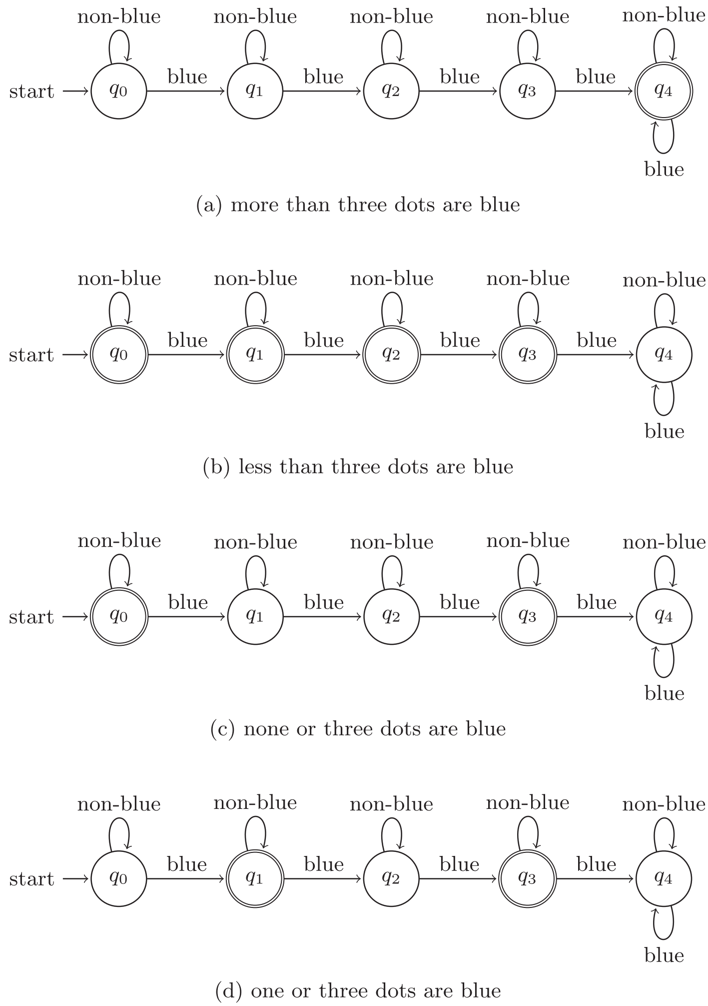


Figure 2 Finite state automata corresponding to the sentences *more than three dots are blue* (Panel 2a), *fewer than four dots are blue* (Panel 2b), *none or three dots are blue* (Panel 2c), and *one or three dots are blue* (Panel 2d). At the beginning of each computation, the automaton is in the start state (indicated by 'start'). Elements of the restriction (i.e. dots) are coded as 0 or 1, here *blue* and *non-blue*. An automaton checks one restrictor element after the other. Depending on whether the dot is blue or not, the automaton performs the respective transition. If the automaton ends up in a accepting (double-circled) state, it returns *true*. If it ends up in a rejecting (single-circled) state, it returns *false*.

and the states are of the same types, two accepting and three rejecting ones, only their ordering is different. Exp. 2 will provide empirical evidence that these two quantifiers clearly differ in processing complexity during verification. This finding is unexpected in the automata theory.

Going beyond simply quantified sentences, the automata account faces another problem. Combining multiple DE quantifiers via quantifier iteration, the overall complexity seems to be affected by the complexity of the individual quantifiers (cf. Experiment 3 below and Bott *et al.* 2013). This intuition can be tested by comparing the following two sentences.

(11) a. Every boy tickled more than two girls.
b. No boy tickled fewer than three boys.

The latter sentence seems intuitively more difficult to understand than the former, but logically they are equivalent. If the observed intuitive difference in quantificational complexity is empirically valid, and Experiment 3 below shows that it in fact is, how can this difference be accounted for in the automata model? The answer is not trivial because the simplest automaton for the iterations of quantifiers in the two sentences turns out to be exactly the same automaton (for details see Steinert-Threlkeld & Icard 2013, p. 169).

We would like to emphasize that this is not intended to imply that the effects reported below provide a priori evidence against the automata account. The first thing to note is that the just given examples only relate to the simplest automata for simple and iterated quantificational sentences. It is conceivable that the present account could be mimicked via semantic automata once other, non-minimal quantifier representations are considered. As usual, we just took the simplest automata for the empty-set and non-empty-set quantifiers that have been proposed in the literature in order to point out that some modification is required to explain the intuitively and experimentally observed differences in quantificational complexity of the examples discussed above. Another thing to note is that the asymmetrical encoding of positive predicate instances seems to be easy to accommodate in the automata framework. Of course only elements from the scope set could be encoded and fed into simplified automata only including the 'positive' transitions (i.e. eliminate all transitions *non-blue* from the automata in Figure 2). In a sense this would mimic the **s-exp** operation introduced above. However, as for the evaluation of empty-set quantifiers it is not obvious how to account for the complexity differences of their evaluation in empty-set situations relative to other situations. Note that independent of the evaluated situation empty-set quantifiers uniformly require an initial accepting state. All three experiments reported below provide clear evidence for empty-set effects. Again, this is not intended to say that the automata theory could not be modified to accommodate these findings. It is just an open question what such a modification – and in particular a solution generalizable to the iteration of quantifiers – would look like.

Another algorithmic account related to the present one is the witness-set-based approach to doubly quantified sentences proposed in Beghelli *et al.* (1997). The paper outlines a procedure using witness sets in order to decide whether it is possible to construct disambiguating situations for doubly quantified sentences consistent with one scope reading but not the other. They proposed the following algorithm for model construction (construction of set diagrams, Beghelli *et al.* 1997, p. 31, their ex. (6)):

(12) To construct a situation that verifies the asymmetrical scope reading $F > G$, pick a witness W_i of the wide scope quantifier F. Using the relation denoted by the predicate, associate with each element of W_i a possibly different witness W_j of the narrow scope quantifier G.

The approach can be exemplified with Beghelli *et al.*'s example *a fireman checks the safety of every building* (ibid., 29). On the linear scope construal, model construction starts with a witness of the wide scope quantifier *a fireman*, for instance, a singleton set comprised of just one fireman. This witness set is connected to the unique witness set of *every building*, i.e. the set of buildings. As for the inverse reading of the sentence, model construction starts with the unique witness set of *every building*, and each element is connected to a potentially different witness of *a fireman*. In other words, firemen may vary with buildings in which case the resulting interpretation is inconsistent with the linear reading.

It is noteworthy that the enterprise of Beghelli *et al.* is orthogonal to the one pursued in the present paper. They propose an algorithm for the *construction* of a disambiguating situation, whereas we propose algorithms for the *interpretation* of a quantified sentence in a given situation. To put it differently, while their account models the interpretation towards a mental model for a scope reading, i.e. the generation of a situation model for a quantified sentence, our account models scope interpretations from a given situation, i.e. the construction and execution of algorithms employed in verification.

Interestingly, Beghelli *et al.* also restricted the proposed algorithm to a subset of quantifying expressions. As it turns out, their proposal only works for UE quantifiers but fails to give correct results for DE and non-monotone quantifiers. The reason is that the latter have upper bounded interpretations which include a maximality requirement. For the present proposal, however, maximality is not an issue since we start with a given situation and can therefore always guarantee maximality of the witnesses. The comparison between the two frameworks suggests that different semantic properties might be decisive for different kinds of interpretation processes. While monotonicity may be essential for comprehension conceived of as model construction, the empty-set property is predicted to be crucial for verification.

3.2 *Experimental findings*

There are actually too many psychological studies on quantifier interpretation and reasoning with quantifiers to allow for a comprehensive review of all the findings reported in the literature (see, among others, Just & Carpenter 1971; Johnson-Laird 1983; Johnson-Laird & Bara 1984; Sanford *et al.* 1994; Tunstall 1998; Paterson *et al.* 2009; Urbach & Kutas 2010; Nieuwland 2016). We therefore focus on psycholinguistic work directly relevant to the current proposal. In particular, we limit the discussion to processing difficulty of single quantified sentences. The only exception is a small excursus on the licensing of complement set anaphora (Sanford *et al.* 1994; Paterson *et al.* 2009). We also neglect studies on the pragmatic interpretation of quantifiers conducted in experimental pragmatics (see, e.g., Gibbs 2017). Furthermore, we will ignore brain imaging studies on the localization of quantifier processing comparing various types of quantifiers in the human brain (the interested reader is referred to McMillan *et al.* 2005; Szymanik 2007; Troiani *et al.* 2009; Wei *et al.* 2012; Olm *et al.* 2014, and the references therein). As it stands, the current proposal is only worked out up to the point that it addresses the algorithmic level and all issues related to implementation (in the sense of Marr 1982) must therefore be left for future research.

3.2.1 *The domain relativizing function of the restrictor argument* Quantificational determiners of natural language are *domain independent* and *conservative*: Their restriction relativizes the domain of quantification (see Section 9 for formal definitions). This universal goes back to Barwise & Cooper (1981) and was further motivated by Keenan & Stavi (1986) who pointed out that it simplifies acquisition because it limits the space of potential quantifier meanings (but, see, Piantadosi *et al.* 2008). On our proposed account, these quantifier universals do not have to be stipulated but rather follow from the theory. What do we know about such properties from experimental research? Recently Hunter & Lidz (2013) presented experimental evidence suggesting that non-conservative quantifiers are in fact impossible to acquire. They tested whether five-year-old children are able to learn new determiners, and constructed a minimal pair consisting of an atrificial conservative and a non-conservative quantifier, which they called *gleeb* and *gleeb'*, respectively.

$$[\![gleeb]\!](A)(B) \Leftrightarrow [\![not\ all]\!](A)(B)$$

$$[\![gleeb']\!](A)(B) \Leftrightarrow [\![not\ all]\!](B)(A)$$

While children were quite successful in learning the meaning of the conservative determiner *gleeb* from only a few instances of determiner uses, none of them was able to figure out the meaning of the non-conservative determiner *gleeb'*. The authors concluded that "this in turn gives us reason to prefer theories of natural language semantics that rule out non-conservative relations as determiner meanings" (Hunter & Lidz 2013, p. 333). We know of no comparable studies having investigated ISOM and EXT, but consider it plausible that this claim also extends to these properties of quantifiers.

Many influential semantic theories (e.g. Barwise & Cooper 1981; Heim & Kratzer 1998) do not exclude non-conservative quantifiers in a principled way. The absence of non-conservative quantifiers must be stipulated as an additional constraint even though conservativity is commonly taken to be a semantic universal (von Fintel & Matthewson 2008). A notable exception is the theory of Keenan & Stavi (1986) who assume a generating class of basic quantificational determiners from which other quantifier meanings are derived via logical connectives. This defines a grammar of determiner meanings which only allows for conservative quantifiers.

The present theory also excludes non-conservative quantifiers right from the outset. The possible witness set quantifiers (as defined in Definition 9.5 in Section 9) are equivalent to generalized quantifiers satisfying CONSERV and EXT (cf. Proposition 9.7). For example, the non-conservative $gleeb'$ cannot be expressed as a w-quantifier because the witness sets of $gleeb'(A)$ would have to contain elements outside of A.[5]

5 Two anonymous reviewers pointed out that this excludes determiner phrases such as *only John* or *every other teacher* from the set of w-quantifiers, which at first sight may seem to be an unwelcome result. In fact, we think that these items must be given another interpretation. *Only* is an operator that associates with focus Rooth (1985, 1992) and therefore not a quantificational determiner. As for *every other teacher*, we analyze *other* as an adjectival modifier of *teacher*, i.e. as part of the restrictor. Evidence for such an analysis comes from the adjectival inflection paradigms of *other* visible in German (cf. Sternefeld 2006) illustrated in (i) and (ii):

(i) a. Jeder betrunkene Lehrer
Every drunk$_{weak}$ teacher
b. Kein betrunkener Lehrer
No drunk$_{strong}$ teacher

(ii) a. Jeder andere Lehrer
Every other$_{weak}$ teacher
b. Kein anderer Lehrer
No other$_{strong}$ teacher

3.2.2 *The time course of quantifier interpretation* Online processing studies of quantifier interpretation suggest that some aspects of quantifier meaning are interpreted immediately but others are only interpreted with some processing delay. This has recently led to a debate about whether quantifiers are interpreted fully incrementally. Their delayed interpretation seems to challenge the incrementality assumption according to which language comprehenders incorporate each word into the linguistic context as soon as it is encountered (Marslen-Wilson 1975; Bott & Sternefeld 2017). In the present theory, different aspects of quantifier interpretation are treated in different interpretation steps, allowing us to separate them for purposes of modeling their real-time interpretation. In the following, we demonstrate that this flexibility is a necessary prerequisite for being able to accommodate a number of findings from the processing literature. In particular, the qualitatively different treatment of a quantifier's restrictor and its scope argument will be shown to be a highly welcome feature for dealing with the realtime interpretation of quantifiers.

Processing the verbal predicate: Processing studies on thematic role assignment show that immediately when the verbal predicate is encountered, arguments are mapped to their respective thematic roles (e.g., Knoeferle *et al.* 2005). In our proposal, thematic roles are hardwired into the respective argument slots of the verbal predicate (cf. Definition 9.25 in Section 9). The encoding of thematic information is thus independent from quantifier interpretation, which involves an extra step, the expansion of the predicate denotation with quantifiers employing the **s-exp** or **c-exp** operation.

Processing the restrictor argument: Whenever the processor encounters a quantificational expression it seems to incrementally restrict its domain. This has been shown in online processing studies on quantifier restriction (Frazier *et al.* 2005; Wijnen & Kaan 2006; Kaan *et al.* 2007; Augurzky *et al.* 2016; Bott *et al.* 2017). A number of studies investigated discourses with bare numerals which either were locally coherent (13-b) with a preceding context sentence (13-a) or not (13-c).

(13) a. Six flowers were put in a vase.
b. Four had a broken stem.
c. Eight had a broken stem.

Eye-movements during reading and reaction times in a stops-making-sense task (Frazier *et al.* 2005; Wijnen & Kaan 2006) revealed processing costs right after encountering the numeral *eight* relative to *four* (but see Kaan *et al.* 2007, for somewhat delayed effects in event-related potentials, ERPs). According to Wijnen & Kaan (2006), this effect reflects disambiguation towards a dispreferred reading where a new set of flowers is added to the discourse in (13-c), while (13-b) is consistent with the preferred subset reading.

This effect can be accounted for in our quantification theory. At the auxiliary *had* it becomes obvious that the phonetically unrealized restrictor argument of the numeral *eight* has to be anaphorically related to a contextually salient set of flowers introduced in the first sentence. If we assume that – in accordance with the preferred subset reading – the determiner *eight* looks for its restriction among the witness sets of *six flowers* and that these consist of sets of exactly six flowers (cf., for instance, Huang *et al.* (2013) on 'exactly readings' of numeral quantifiers), then a conflict emerges at this point during discourse processing: No matter which witness set X of *six flowers* is chosen as restriction for *eight* and what the upcoming sentence predicate will be, expansion with the quantifier *eight*(X) fails. A way out to maintain discourse coherence is to accommodate a larger set of flowers.

According to this interpretation, the set of flowers in the vase and those with a broken stem form distinct subsets – in accord with our intuitions about the meaning of these discourses. Adding the rather uncontroversial assumption that accommodation incurs processing cost explains why (13-c) is more costly than (13-b).

Further evidence for the incremental interpretation of quantificational restriction comes from a recent ERP study by Augurzky *et al.* (2016) and behavioral experiments by Bott *et al.* (2017) investigating the processing of quantified questions with extraposed restrictive relative clauses such as *Are all triangles* <u>*blue*</u> *that are in the circle?* (lit. from German). At the underlined critical colour adjective, that is even before the presentation of the extraposed relative clause, ERPs already differed as a function of a previously presented visual context either suggesting a subset reading of *circle* (e.g., blue dots inside and red dots outside the circle), or no subset reading (all dots of the same colour). Behavioral experiments within the same design Bott *et al.* (2017) show that a full-fledged answer to the question has already been motorically prepared mid sentence at the adjective, and that the processor incrementally takes into account the properties encoded in the context representation. These findings provide further evidence for the incremental interpretation of quantificational restriction.

In the present account interpretation of quantifiers can only start after having determined their restrictor argument. It is therefore highly expected that the restrictor set is accessed right away during realtime interpretation.

The scope argument of simply quantified sentences: The proposed division into different interpretation steps is highly welcome once we consider studies investigating the time course of the 'second argument' of quantificational determiners, i.e. their nuclear scope. A number of ERP studies have addressed incremental interpretation of quantifiers by investigating simply quantified sentences that do not correspond to our commonsense beliefs about the actual world using the N400 component of the ERP as dependent measure. The N400, a negative deflection in the ERP waveform (i.e. stimulus induced electrical activity as part of the EEG measured at the scalp) peaking around 400 ms post onset of the critical stimuli, is commonly thought to reflect the degree of how well the incoming stimulus fits into the representation built up so far (see Kutas & Federmeier 2011, for a review). Claims from the literature range from non-incremental over partially incremental to fully incremental views on quantifier interpretation (Kounios & Holcomb 1992; Urbach & Kutas 2010; Urbach *et al.* 2015; Freunberger & Nieuwland 2016; Nieuwland 2016).[6] In the first study on quantifier comprehension, Kounios & Holcomb (1992) manipulated quantifier type (*all* vs. *some* vs. *no*) affecting the truth value of sentences such as *all/some/no rubies are gems*. They did not find an effect of quantifier type on the N400 elicited by the last word of the sentence, even though truth-value judgments indicated that participants had interpreted the sentences correctly. The authors concluded that quantifier interpretation is delayed and that the N400 does not reflect late compositional interpretation processes involving quantifiers. This was taken up by Urbach & Kutas (2010) manipulating quantifier type ('positive' vs. 'negative'; e.g., e.g., *most* vs. *few*) fully crossed with a truth-value manipulation

6 As noted above, we are limiting the discussion to 'semantic' aspects of quantifier interpretation. Another, completely orthogonal issue is the relative timing of pragmatic aspects of quantifier meaning (see e.g. the discussion concerning the time course of scalar implicatures in Huang & Snedeker 2009 and Grodner *et al.* 2013, among others).

(*most/few farmers grow crops/worms*). End of sentence plausibility ratings showed the expected cross-over interaction, but this was not the case in the N400 component. The type of quantifier modulated the N400 in the expected direction, but the pattern was not completely reversed in the negative quantifier conditions. This led the authors to the conclusion that quantifier interpretation proceeds 'partially incremental' with delayed processing of negative quantifiers. Nieuwland (2016) employed the same design as Urbach & Kutas (2010) but additionally controlled for the expectedness of the critical word. For items with low cloze probability effects were similar to Urbach & Kutas's study, but for items with a high cloze probability of the critical word, N400 effects were as predicted by the incrementality hypothesis. He concluded that quantifier interpretation can be fully incremental but only if there is sufficiently strong guidance by world-knowledge about typical situations for developing predictions about how the quantified statement will continue. The recent studies by Urbach *et al.* (2015) and Freunberger & Nieuwland (2016) point in a similar direction. If quantified sentences are embedded in a real-world context Urbach *et al.* (2015) or their interpretation is supported by prosodic information in auditory presentation Freunberger & Nieuwland (2016), incremental N400 effects are observed for both positive and negative quantifiers. Urbach *et al.* (2015) furthermore show that the occurrence of a fully crossed interaction in addition depends on the experimental task. The interaction was only observed in a reading-for-comprehension task, but not in a plausibility-rating task.

To summarize, 'negative' quantifiers in contrast to 'positive' quantifiers sometimes seem to be interpreted in a delayed fashion and seem to require additional contextual support. Relating these findings to the proposed theory, it is worth noting that all 'positive' quantifiers tested belong to the class of non-empty-set quantifiers, while all 'negative' quantifiers were empty-set quantifiers. According to our proposal, the 'positive' quantifiers in these sentences could thus be evaluated using the **s-exp** algorithm, but the 'negative' ones required the more complex **c-exp** algorithm. This difference in complexity, both in terms of encoding negative predicate instances as well as in terms of the additional clauses required for successful quantifier expansion, provides us with an explanation – not the only conceivable one, of course – for the observed delay in processing effects. Furthermore, it is plausible to assume that if the context makes the negative predicate instances salient, the processing of 'negative' quantifiers should be sped up, which would in turn account for the observed contextual facilitation effects.

Relative scope in multiply quantified sentences: Turning to the time course of scope computation in doubly quantified sentences, Bott & Schlotterbeck (2015) investigated the incremental nature of processing costs of German doubly quantified sentences involving scope reconstruction (see the references therein for psycholinguistic studies on the processing of quantifier scope ambiguities). They addressed the question whether the relative scope of quantifiers is resolved incrementally or in a more global fashion at the end of a minimal complete sentence. In one self-paced reading and one eyetracking during reading experiment they showed that scope reconstruction was indeed delayed until the end of the sentence, or rather until the point when readers encountered both quantifiers and the verbal predicate (see also Radó & Bott (2012) for another study on relative quantifier scope pointing in the same direction and the discussion in Sanford & Sturt (2002)). They presented sentences such as (14), in which binding of the possessive pronoun in the first quantifier enforces scope inversion, and investigated when during reading such sentences scope inversion costs would show up.

(14) Jeden seiner Schüler hat genau ein Lehrer voller Wohlwollen gelobt.
Each of his pupils has exactly one teacher full of goodwill praised.
'Exactly one teacher praised each of his pupils full of goodwill.'

Their eyetracking during reading experiment revealed clear evidence for a scope inversion effect. This effect, however, was restricted to eyetracking measures reflecting the rereading of these sentences whereas during first-pass reading the scope inversion sentences were indistinguishable from control sentences with a surface scope interpretation. These results are fully consistent with the proposed theory of quantification, in which it was implicitly assumed that the evaluation of multiply quantified sentences depends on the complete sentence. Note that the execution of the proposed algorithms for surface scope in doubly quantified sentences actually started by expanding the verbal predicate with the linearly second and not the first quantifier.

To sum up, the relative difference in the timing of thematic interpretation, the interpretation of quantifier restriction, and scope interpretation in simply and multiply quantified sentences can be accommodated in our theory by the proposed separation of these three aspects of quantifier interpretation.[7] Thus, the proposed theory is intended to be flexible enough to be able to model a general separation of various aspects of semantic interpretation. We would like to point out that this does not lead to a complete arbitrariness order of processing steps. While thematic roles can be assigned independently of scope, it is not possible to interpret scope without having assigned quantifiers their thematic roles first.

3.2.3 *Inherent complexity differences between quantifiers* Another desideratum for any cognitively realistic theory of quantification is that it should be able to account for inherent differences in the semantic complexity between different kinds of quantifiers.

Monotonicity: One semantic property that has repeatedly been shown to affect quantificational complexity is the monotonicity of quantifiers (in their right argument). Already Barwise & Cooper (1981) hypothesized that DE quantifiers should be more difficult to verify than UE ones. This hypothesis has been confirmed by a number of experimental studies. Above, we already discussed ERP evidence that the polarity or monotonicity of quantifiers affects their time course of interpretation. Here, we focus on studies that investigated whether monotonicity affects how difficult it is to evaluate quantified sentences against pictures.

Just & Carpenter (1971) investigated the closely related property of *affirmativity* and provided experimental evidence that negative quantifiers take longer to process than affirmative quantifiers.[8] Their picture verification experiments included simply quantified

7 In Definition 9.25 in Section 9, we introduce Cooper storage to assign thematic roles to DPs independently of quantifier scope which will only be introduced in the subsequent definitions.

8 Although being closely related with each other, negativity must not be identified with downward monotonicity. Negativity is a linguistic concept defined via question tags and the licensing of negative polarity items such as *ever*. By contrast, monotonicity is a logical/mathematical concept. The class of monotone decreasing quantifiers includes affirmative quantifiers as, for instance, illustrated by the following two examples with monotone decreasing quantifiers:

(i) a. At most one third of all linguistic students ever visited a formal semantics class.
b. *Not all linguistic students ever visited a formal semantics class.

sentences such as (15-a–d) and they measured reaction times (RTs) for the two types of quantifiers. RTs were greater for negative sentences than they were for affirmative sentences.

(15) a. Many of the dots are red. (affirmative/increasing)
b. Few of the dots are red. (negative/decreasing)
c. The majority of the dots is red. (affirmative/increasing)
d. The minority of the dots is red. (negative/decreasing)

Effects of monotonicity on RTs in sentence-picture verification were later replicated by Koster-Moeller *et al.* (2008), Geurts *et al.* (2010), Szymanik & Zajenkowski (2013), Deschamps *et al.* (2015) and Schlotterbeck (2017). All of these studies reported longer judgment RTs for monotone decreasing quantifiers than for increasing ones. To give a concrete example Deschamps *et al.* tested the effects of linguistic instructions compared to non-linguistic instructions in a verification task manipulating the proportions of yellow and blue dots as well as the monotonicity of the following three types of quantifiers, cf. (16):

(16) a. More/Less than half of the dots are blue/yellow.
b. Many/Few of the dots are yellow/blue.
c. There are more/fewer blue/yellow dots than yellow/blue dots.

Their experiments revealed robust main effects of monotonicity across quantifier types in the absence of interactions with the different proportions. For all three quantifier types monotone decreasing quantifiers led to more errors and longer judgment RTs than monotone increasing quantifiers. This effect was absent in the non-linguistic instructions showing a yellow and a blue square with an inequality sign in between them. Deschamps *et al.* (2015) interpreted this complex pattern of results to suggest that monotone decreasing quantifiers are representationally more complex than monotone increasing ones.

According to the present proposal the observed main effect of monotonicity is not surprising. All DE quantifiers have the empty set among their witness sets and thus sharply contrast with AE non-empty-set quantifiers as they were tested in the reviewed studies. Accordingly, all downward entailing quantifiers require **c-exp**, whereas the tested increasing quantifiers could be evaluated with the simpler algorithm **s-exp**.

Another finding that has been repeatedly observed in quantifier verification experiments is an interaction between MONOTONICITY and TRUTH VALUE (Just & Carpenter 1971; Koster-Moeller *et al.* 2008; Szymanik & Zajenkowski 2013). But the picture that emerges from these data is not a simple one: Whether such an interaction is indeed present in a given experiment and which form it takes crucially depends on multiple factors, like the order in which linguistic and visual stimuli are presented (as shown by Just & Carpenter), the experimental task (compare, e.g., the study by Koster-Moeller *et al.* with that of Szymanik & Zajenkowski) or the type of quantifier (e.g. proportional vs. cardinal, as investigated by Szymanik & Zajenkowski, but see also differences between the quantifier classes of Just & Carpenter). A detailed discussion of these subtle differences goes beyond the scope of the present paper. Nevertheless, we would like to mention that the present proposal is generally compatible with various explanations that have been given for such interaction effects. This is because we are not suggesting to alter the standardly assumed truth conditions or semantic properties of quantified statements. For example, it still holds in our proposal that during the evaluation of UE quantifiers a *false*-judgment may have to be revised if additional information is taken into account, whereas for DE quantifiers it is the *true*-judgment that always leaves some uncertainty. For a discussion of these issues from the perspective of the semantic automata account the interested reader is referred to Szymanik (2016) who makes fine-grained distinctions between different 'stages' of the verification process.

3.2.4 *Other properties* Turning to properties different than monotonicity but also influencing quantificational complexity we would like to point out some of the limits of the present account. One of the central points in the mentioned study by Geurts *et al.* (2010) were processing differences between comparative quantifiers (*more/fewer than n*) and superlative quantifiers (*at least/at most n*) and, more generally, processing differences between so-called class A and class B modified numerals Nouwen (2010), of which the comparative-superlative distinction is a special case. As it stands, our theory does not account for these differences. To do so we would at least have to model ignorance inferences since this is what seems to set these quantifier types apart (among others, see e.g. Geurts & Nouwen 2007; Büring 2008; Cummins & Katsos 2010; Coppock & Brochhagen 2013; Schwarz 2013, 2016; Spychalska 2018).

Although we have seen repercussions of the distinction between proportional and other types of quantifiers in Section 2, we also have nothing new to say about the processing difficulty that has been attributed specifically to proportional quantifiers McMillan *et al.* (2005); Szymanik & Zajenkowski (2010, 2013). We would like to suggest that these differences should ultimately be dealt with in spelling out a model of how the cognitive system identifies the witness sets of a quantifier (for a discussion of this problem cf., e.g. Robaldo *et al.* 2014). We leave this and related issues for future research.

So far we have looked at the interpretation of the quantificational statement itself. Even though our theory is not a discourse theory, it could be seen to make some predictions for discourse interpretation. To these we now turn.

3.2.5 *Quantifiers in discourse: Licensing of complement set anaphora* Psycholinguistic research has demonstrated striking differences between quantifiers in terms of plural anaphora they can license. Again, it turns out that monotone increasing quantifiers behave differently from monotone decreasing ones – or at least this appears to be a good first approximation (cf. Moxey *et al.* 2001; Nouwen 2003).

(17) a. A few of the fans went to the match. Their presence ...
b. # A few of the fans went to the match. Their absence ...
c. Few of the fans went to the match. Their presence ...
d. Few of the fans went to the match. Their absence ...

The example shows that *a few of the fans* only gives rise to what has been termed reference to the *reference set* (e.g., the fans that went to the match) whereas *few of the fans* also allows for what has been called reference to the *complement set* (e.g., those that did not go). An influence of quantifier polarity has been empirically confirmed in a number of production and comprehension experiments. These experiments employed a variety of methods including eyetracking during reading and the recording of ERPs during story comprehension (Moxey & Sanford 1987, 1993; Sanford *et al.* 1994; Sanford *et al.* 1996; Paterson *et al.* 1998; Moxey *et al.* 2001; Sanford *et al.* 2001; Moxey 2006; Sanford *et al.* 2007; Paterson *et al.* 2009; Filik *et al.* 2011).

The present proposal is clearly not a discourse theory and it does not model anaphoric relations (see, e.g., Nouwen 2003, for a comprehensive theory). However, considering the four cases in (17), our proposal may provide an explanation of what may be called the semantic licensing conditions of complement-set anaphora: The first sentence of (17-a,b) containing the non-empty-set quantifier *a few of the fans* can be evaluated with the **s-exp** algorithm. By contrast, an empty-set quantifier such as *few of the fans*, requires the **c-exp** expansion operation. Only a small step is missing from here to an explanation of how the polarity of the quantifier determines (un-)availability of complement set reference.

The **s-exp** procedure only evaluates positive predicate instances and it depends on the individuals in the reference set but not those in the complement set. Thus, the **s-exp** procedure ignores the complement set and it is therefore not surprising that this set is not available for anaphoric reference. This is different for quantifiers that require the **c-exp** algorithm. Above, we have shown that the positive predicate instances as well as the negative predicate instances are important for correct evaluation of empty-set quantifiers. In consequence, we propose that negative predicate instances – in addition to positive ones – are explicitly represented only when processing empty-set quantifiers and that this is in turn a necessary precondition for anaphoric reference.[9]

3.2.6 *Summary* Our proposal naturally connects to a large number of findings from the psycholinguistic literature that have not been dealt with in a unified account before. The proposed theory explains the domain-relativizing function of the restrictor argument of natural language quantifiers. It accounts for enhanced processing difficulty of DE as compared to UE quantifiers. It separates different aspects of quantifier interpretation from each other and treats them independently from each other. This turns out to be highly welcome for modeling the time-course of quantifier interpretation with the observed incremental and non-incremental effects. Finally, the present proposal can be used to spell out licensing conditions of anaphoric expressions that may explain some of the behaviour of complement set anaphora.

9 There are also some more subtle observations and the proposed theory does not account for these. In particular, we would like to briefly mention four points. Firstly, psycholinguists have shown that the different anaphoric possibilities interact with discourse expectations. These are beyond the scope of the present proposal (cf. Moxey 2006). Secondly, Nouwen (2003) shows that complement-set anaphora are limited to proportional, decreasing quantifiers. By contrast, complement-set reference is generally ruled out for cardinal quantifiers irrespective of their monotonicity, as illustrated by the following example.

(i) Fewer than five students went to the party. Their presence/#absence . . .

Thirdly, Nouwen (2003) also points out that reference to the complement set seems to be only possible if an anaphoric link to the reference set would result in an infelicitous interpretation, as in (17-d). Finally, he proposes that the complement set has a very different discourse status than the reference set: It can be inferred from the discourse, but it is not explicitly given by the discourse. According to Nouwen (2003), reference to the complement comes about by the inference that there is a – necessarily non-empty – salient set that is constructed by subtracting the scope set from the restrictor set (but see (Filik *et al.* 2011)). All these points go well beyond the scope of the present proposal. Nevertheless, we would like to point out that the conception of quantification presented here is consistent with the the idea that the complement set comes about by a potentially costly inference.

In addition, the proposed theory makes predictions that have not been tested experimentally yet. To our knowledge, in psycholinguistic research the empty-set property has so far not been addressed as a source of quantificational complexity. We will now turn to the hypotheses and predictions tested in the experiments reported below.

4 HYPOTHESES

In the following, we derive a set of hypotheses concerning processing difficulty of quantifier interpretation. The first concerns verification of empty-set quantifiers in empty-set situations. As motivated above in Sections 1.1 and 2, the derivation of a positive conclusion from negative predicate instances should be particularly costly. This is what is implemented in clause 3 of the **c-exp** operation, licensing an inference from negative information, i.e. the absence of positive predicate instances, to a positive evaluation of an empty-set quantifier..

Hypothesis 4.1 (Empty-Set Effects). *In empty-set situations, the verification of empty-set quantifiers leads to enhanced difficulty relative to situations with positive predicate instances. No such effect occurs with non-empty-set quantifiers.*

The next two hypotheses concern processing difficulty of empty-set quantifiers that is independent of the context against which they are evaluated. They correspond to two theoretical alternatives as to how the **s-exp** versus **c-exp** operations may be used to evaluate sentences with more than one quantifier (see Section 9 for technical details). For sentences with just one quantifier they are equivalent, but they diverge for sentences with more than one quantifier.

In principle they apply to both verification and comprehension but a comment on the role of comprehension as opposed to verification data may be in order here. First and foremost the proposed theory concerns *verification* algorithms constructed during quantifier interpretation. It therefore directly relates to verification data. With regard to comprehension the relationship between theory and processing data is far more indirect. We can think of a number of alternative ways in which verification procedures may be prepared or constructed during online processing, ranging from fully specified algorithms, e.g. in the form of decision trees for complete quantified sentences to simple lists of **s-exp** or **c-exp** clauses for the individual quantifiers. We may even consider the possibility of underspecified semantic representations (see, e.g., (Reyle 1993; Sanford & Sturt 2002)). The comprehension data reported below are therefore interesting in their own right but their relation to the proposed theory is far more indirect than that of the verification data. As we move along the experiments, it will become obvious that the comprehension data reflect only some of the effects present in the verification data. Other effects, however, seem to be absent during comprehension. The direct comparison of the two types of data therefore grants us some access to the nature of the representations constructed during the incremental interpretation of quantified sentences.

According to the first hypothesis the expansion operation – either **s-exp** or **c-exp** – is determined in a wholistic fashion for the whole sentence. This means that a single occurrence of an empty-set quantifier is enough to enforce evaluation of all quantifiers in the sentence with the **c-exp** operation. Or, to put it differently, **s-exp** is reserved to the evaluation of quantified sentences without any empty-set quantifiers. This predicts increased difficulty during verification and comprehension as soon as one of the quantifiers has the empty set as a witness set. Simple expansion uses only one rule, whereas additional rules are required for the complex expansion operation. In addition, it requires the encoding and manipulation of negative predicate instances. These two aspects are expected to lead to enhanced processing difficulty.

Hypothesis 4.2 (General difficulty of empty-set quantifiers, alternative A). *In quantified sentences with at least one empty-set quantifier, as opposed to sentences without any empty-set quantifiers, processing difficulty during verification (and comprehension) is enhanced due to inherent complexity differences between expansion operations. In multiply quantified sentences, one empty-set quantifier is enough to trigger the need for the complex expansion operation. Thus, adding other empty-set quantifiers does not increase processing load (relative to adding a non-empty-set quantifier).*

Alternatively we may assume that the comprehension and verification of quantified sentences involves the simplest sequence of individual expansion operations – either **s-exp** or **c-exp** – for evaluating each individual quantifier in the sentence. In contrast to the first theoretical alternative stated in Hypothesis 4.2 this theoretical alternative predicts cumulative effects of empty-set quantifiers. For each empty-set quantifier the complex expansion has to be prepared during comprehension and executed during verification. Therefore, each individual empty-set quantifier should additively contribute to overall processing difficulty. The two alternative hypotheses have been directly contrasted with each other in example (10).

Hypothesis 4.3 (General difficulty of empty-set quantifiers, alternative B). During verification (and comprehension) of quantified sentences each empty-set quantifier (as compared to non-empty-set quantifiers) increases processing load in a cumulative fashion.

5 EXPERIMENT 1 – QUANTIFICATIONAL COMPLEXITY OF SIMPLY QUANTIFIED SENTENCES WITH MODIFIED NUMERALS

In the first experiment the effects of quantificational complexity due to monotone increasing, decreasing, and non-monotone quantifiers were investigated in simply quantified sentences. Quantifiers were chosen in such a way that only the monotone decreasing quantifier was an empty-set quantifier. According to the present approach, only the monotone decreasing quantifier *fewer than five dots* should therefore lead to processing difficulty whereas the non-monotone quantifier *exactly five dots* should pattern with the monotone increasing quantifier *more than five dots*. This prediction sets our proposal apart from Beghelli *et al.* (1997) who group non-monotone quantifiers together with monotone decreasing ones.

Furthermore, the present experiment compared quantifier verification in empty-set situations to the verification of non-empty-set situations (testing Hypothesis 4.1). A prediction unique to the present account is that verification difficulty of the empty-set quantifier *less than five dots* should be increased in an empty-set situation relative to a non-empty-set situation – an effect predicted to be absent for the non-empty-set quantifiers *exactly five dots* and *more than five dots*.

As in all experiments to follow comprehension as well as verification data were logged in each trial. Participants first read a quantified sentence while measuring their reading times. After reading they went on to a verification stage in which they had to provide a truth-value judgment for a picture presented on a separate screen. As for the verification stage, judgment RTs and the proportions of errors were logged and subjected to statistical analyses.

Both the quantifier and the number of target objects in the scope set were manipulated to see whether quantifiers differed in their inherent complexity and whether any empty-set effects were present. The experiment was designed in such a way to be able to rule out an alternative pragmatic explanation related to a potential scalar implicature of what might look like an empty-set effect. Specific predictions will be outlined in Section 5.1.4 after having presented the experimental methods.

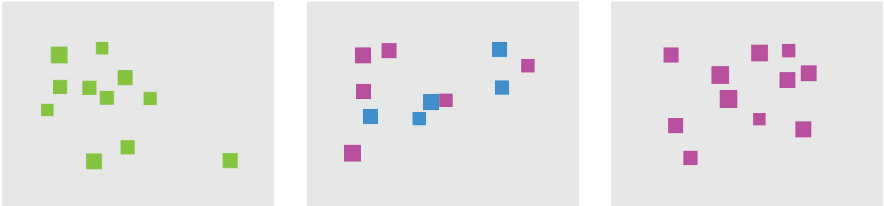
5.1 *Methods*

5.1.1 *Participants* 48 students (mean age 24.3 years, range 18-36 years, 42 female) from the University of Tübingen participated in the experiment for payment of €5. An experimental session took no longer than 15 minutes. 16 participants were randomly assigned to each list.

5.1.2 *Materials* The experiment employed German versions of the sentences illustrated in (18). Slashes indicate segmentation for self-paced reading. *More than five* (*mehr als fünf*) is monotone increasing, *fewer than five* (*weniger als fünf*) is monotone decreasing and *exactly five* (*genau fünf*) is non-monotone. Sixteen experimental sentences were constructed in three quantifier conditions varying geometrical forms (four forms) and colors (four colors) mentioned in the sentence.

(18) a. Mehr als fünf Punkte / sind blau.
More than five dots / are blue.
b. Weniger als fünf Punkte / sind blau.
Less than five dots / are blue.
c. Genau fünf Punkte / sind blau.
Exactly five dots / are blue.

The sentences were paired with pictures showing 11 objects of the geometrical form encoded by the restrictor. We manipulated the number of objects that were of the target color mentioned in the sentence. Henceforth, we will refer to the resulting models as 0-, 1-, 2-, 3-, . . ., 10-, 11-models (→ a total of 12 types of models). Counterbalancing procedures were used to create 144 unique pictures which showed the objects in randomized positions with all combinations of colors appearing equally often across the experiment. Sample pictures are shown in Figure 3.



(a) 0-model (b) 6-model (c) 11-model

Figure 3 Sample pictures from Experiment 1 for *a) more than b) fewer than c) exactly five squares are pink*. *0-model* = zero target objects, *6-model* = six target objects, *11-model* = eleven target objects.

Each experimental sentence was tested in all three quantifier conditions with exactly the same picture and a latin square was employed to distribute experimental sentences to three lists such that each participant saw each picture only once with one out of three quantifiers, while getting to see each quantifier/model combination ($3 \times 12 = 36$ combinations) four times. This way, each quantified sentence was tested with exactly the same picture materials. Since only the 0-, 1- and 10-, 11-models are relevant for testing the predictions of the proposed theory, the 2-, 3-, ..., 9-model conditions will be treated as fillers.

5.1.3 *Procedure* First participants read a quantified statement which was presented self-paced with moving window presentation Haberlandt (1994). They had to press the space bar on the computer keyboard to get to see the next sentence region while all characters outside this region were replaced by dashes. After the last sentence region, the sentence automatically disappeared and was replaced by a new screen showing the picture in centered position. A truth-value judgment had to be given within a time limit of five seconds. Participants were instructed to decide as quickly as possible whether the sentence was true or false by pressing the "yes, true" or "no, false" button with their left or right index finger. The response key mapping was counterbalanced across participants. For half of the participants *true* was mapped to their left index finger and for the other half it was mapped to their right index finger. The experiment consisted of a single block with individually randomized order of experimental trials. At the beginning of this block, participants completed three extra trials as a practice to get used to the task.

5.1.4 *Predictions* A number of factors can be expected to affect difficulty of the verification task. In the following we would like to focus entirely on predicted effects related to the processing of the empty-set quantifier relative to the other two. For this reason, we will only discuss the relevant conditions, that is 0- and 1-models as well as 10- and 11-models for all three quantifiers, and treat the other conditions as fillers.

Crucially, our theory predicts an empty-set effect, i.e. difficulty in evaluating 0-models relative to 1-models exclusively for *fewer than five* because neither *more than five* nor *exactly five* are empty-set quantifiers. Recall that the evaluation of an empty-quantifier in a 0-model is the only condition that requires an inference from the absence of objects to the truth of the quantified statement. Such situations are dealt with by the third clause of the **c-exp** algorithm. In Section 4 (Hypothesis 4.1), we hypothesized that the application of this clause should be harder than any of the other clauses because it involves the encoding of negative information as well as polarity reversal. By contrast, for the evaluation of the 1-model only a cardinality judgment of the set of positive predicate instances has to be made applying the second clause of the **c-exp** verification procedure. It just has to be determined that the set of predicate instances has cardinality 1 and that $1 < 5$. On top of this effect we expected that the empty-set quantifier *fewer than five* leads to enhanced difficulty independently of the presented picture because it should translate into the more complex **c-exp** expansion algorithm. The latter prediction applies to its comprehension as well as to its verification.

In addition to empty-set effects we tested for potential pragmatic effects having to do with underinformative quantificational statements. What might look like a prima facie empty-set effect might also reflect a pragmatic effect. This is because for sentences with the quantifier *fewer than five* presented in the context of a 0-model sentences with the quantifiers *no* might be more informative competitors. For example, *no dot is blue* would be a more informative alternative to the sentence in (18-b) if presented in the context of zero blue dots.[10]

It can thus be expected to violate Grice's maxim of quantity and therefore give rise to a scalar implicature *fewer than five but not none* even though these quantifiers are not among the prototypical examples discussed in the pragmatic literature (Horn 1968; Grice 1991; Levinson 2000; Geurts 2010). To decide whether such a pragmatic effect was present in the experiment we analyzed *more than five* sentences in 11-models. In this type of model (11 out of 11 objects are target objects) *more than five* sentences are less informative than universally quantified sentences such as *all dots are blue*. Comparing 0-models vs. 1-models for *fewer than five* with 11-models vs. 10-models for *more than five* should put us in the position to investigate whether apparent empty-set effects are actually scalar-implicature effects in disguise.

Concerning reading times, effects of quantificational complexity were expected as soon as participants enter the region with the verbal predicate *are blue* (*sind blau*), which was held constant across sentence conditions. This expectation was based on the fact that this is the earliest point at which a verification algorithm can be fully specified. Because only *fewer than five dots* is an empty-set quantifier, we expected longer reading times for this quantifier than for the other two.

5.1.5 *Statistical analysis* Reading times, judgment RTs, and error rates were analyzed as follows. The reading times and judgment RTs were corrected for outliers by removing all RTs below 100ms and RTs more than 2.5 standard deviations above the mean RT for any given participant. This affected 3.0% of the reading time data and 3.3% of the verification data. For descriptive statistics we will report the aggregated data over participants. Repeated measures analyses of variance (ANOVAs) were computed analyzing the reading times of the final sentence region containing the nuclear scope contingent on QUANTIFIER.[11] Due to the lexical differences in the first part of the sentence (containing the quantifier), no inferential statistics were computed for this region. Judgment RTs were analyzed in two ANOVAs.[12] To test for pragmatic effects of scalar alternatives a 2×3 repeated measures analysis with the factors MODEL and QUANTIFIER was computed. The analysis compared the 10- vs. 11-models to test for any difficulty of *more than* in the 11-models. Another ANOVA was computed to confirm that there was a reliable empty-set effect. For this purpose another 2×3 ANOVA including the within factors MODEL and QUANTIFIER was computed on the judgment RTs in the 0- and 1-models. Errors were subjected to logit mixed-effect model analyses. Since models with interactions in the random slopes failed to converge, interaction terms were omitted from the random effects structure and only the main effects of QUANTIFIER and MODEL were included in the slopes. Here and also in the two experiments reported below, an α-level of .05 was chosen and we report the significance levels $p < .05$, $p < .01$ and $p < .001$.

10 We would like to thank the audience at the University of Konstanz to whom we presented an earlier draft of this paper and in particular Maribel Romero (p.c.) for pointing out this alternative explanation.

11 Because the experimental items were identical except for the nouns and the color adjective we do not report any by items (F_2) analyses for the present experiment Clark (1973). The by-participants analysis (F_1) will be simply referred to by F without a subscript.

12 Additional analyses were run on judgment RTs only including the RTs of correct judgments. Since the pattern of effects was the same for both sets of analyses, we refrain from reporting them here.

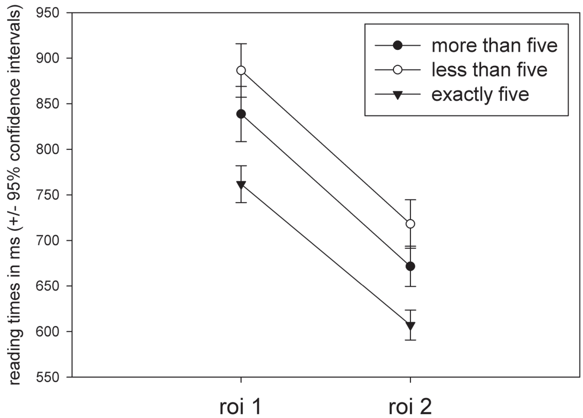


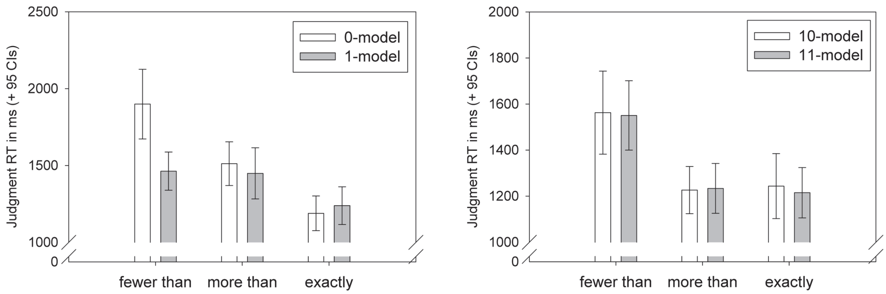
Figure 4 Mean reading times in Exp. 1 depending on quantifier type (+ 95% confidence intervals, computed on the basis by-participant means). roi 1: quantifier + restriction, roi 2: nuclear scope.

5.2 *Results and Discussion*

Figure 4 shows the mean reading times for all three quantifier types. Across both sentence regions, participants spent longer reading sentences with *fewer than five* than they spent on reading sentences with *more than five*, which in turn took longer to read than sentences with *exactly five*. At the sentence final region this difference was reflected by a significant main effect of QUANTIFIER ($F(2, 94) = 20.87, p < .001$). Paired t-tests further confirmed that both the reading time differences between *fewer than five* and *more than five* ($t(47) = 3.58, p < .001$) and between *more than five* and *exactly five* ($t(47) = 4.87, p < .001$) were significant. This result supports our hypothesis that the empty-set quantifier *fewer than five* would give rise to a more complex algorithm (**c-exp**) than the other two quantifiers (cf. Section 4). Interestingly, the non-monotone determiner *exactly five* was read even faster than the monotone increasing determiner *more than five*. This lends support to our proposal that non-monotone quantifiers do not pattern together with monotone decreasing ones.[13]

Figure 5 presents the mean judgment times for the three quantifiers depending on the type of model. The mean percentage of correct judgments are summarized in Table 1. The analysis of proportions of errors showed that the evaluation of *fewer than five* in a 0-model was the only condition that led to a substantial amount of errors (25.0% “no, false” judgments) while each of the other conditions received less than 10% errors. In a logit mixed effects model analysis including QUANTIFIER and MODEL (two levels: *0-* vs. *1-model*) this difference was reflected by the predicted interaction (estimate= 4.57, $z = 2.39$, $p < .05$) and a main effect of QUANTIFIER (*fewer than five* vs. the other two quantifiers: estimate= −6.32, $z = -3.49$, $p < .001$). The same pattern showed up in the verification times. The falsification of *0-models* after *fewer than five* took on average 1900ms relative to 1463ms in *1-models*. For the other two quantifiers RTs were roughly

13 To test whether the observed effect could be due to spillover from the first region including lexical differences, an additional linear mixed effects model analysis was conducted including the RTs of the first region as predictors for the RTs of the second. This analysis also revealed a significant fixed effect of QUANTIFIER with the same contrasts (*estimate* = 138.7, $t = 4.48$). We would like to thank an anonymous reviewer for pointing out the possibility of a spillover effect.



(a) *Empty-set analysis* (b) *Pragmatic analysis*

Figure 5 Mean verification times in Exp. 1 contingent on the number of target objects in the model and quantifier type (+ 95% confidence intervals). Panel (a) presents the mean judgment RTs taken into account for the empty-set effect analysis. Panel (b) presents the mean judgment RTs taken into account for the analysis of a potential pragmatic effect. Note: 0-models are situations in which none of the restrictor elements was part of the scope set; 11-models are situations in which all of them were in the scope set.

Table 1 Mean proportions of correct answers in % (standard deviations) contingent on QUANTIFIER and MODEL in Exp. 1.

		Determiner	
	Fewer than five	*Exactly five*	*More than five*
0-model	75.0 (33.0)	99.5 (3.6)	97.4 (7.7)
1-model	94.8 (10.3)	98.4 (6.1)	97.4 (9.3)
10-model	95.8 (9.4)	98.4 (6.1)	95.8 (14.0)
11-model	97.9 (7.0)	99.5 (3.6)	96.4 (8.9)

the same (*exactly five*: 1188ms vs. 1239ms; *more than five*: 1512ms vs. 1449ms). In the ANOVA on verification RTs this was reflected by significant main effects of QUANTIFIER ($F(2, 94) = 28.64, p < .001$), MODEL ($F(1, 47) = 9.32, p < .001$), and most crucially the predicted interaction between QUANTIFIER and MODEL ($F(2, 94) = 10.13, p < .001$). Planned comparisons revealed a significant RT difference between *0-* and *1-models* when evaluating *fewer than five* ($t(47) = 4.21, p < .001$), but not in evaluating the other two quantifiers (both $|t| < 1$).

To rule out the pragmatic alternative explanation of the observed empty-set effects, the three quantifiers were compared in the verification of *10-* vs. *11-models*. The latter were models where all objects were of the target color and *more than five* sentences might therefore be perceived as underinformative. By contrast, in the *10-models* pragmatic competition should not be at issue because 10 out of 11 objects does not mark the end of the scale. The analysis of acceptance rates and verification times showed no indication of a pragmatic effect. Participants accepted *more than five* sentences in *11-models* 96.4% of the time and in *10-models* (95.8%). The analysis of verification RTs further corroborated this lack of pragmatic effect. An ANOVA revealed only a reliable main effect of QUANTIFIER

($F(2, 94) = 25.76, p < .001$), but neither a significant main effect of MODEL nor a reliable MODEL×QUANTIFIER interaction ($F < 1$).[14]

To summarize, the presented data are fully consistent with the proposed quantification theory. The reading times support the hypothesis that empty-set quantifiers are generally difficult: the specification of an algorithm for an empty-set quantifier is inherently more complex than the comprehension of non empty-set quantifiers such as *exactly five* and *more than five*. Recall that this follows from the number of rules required for the S-EXP and C-EXP algorithms. The hypothesis that this difficulty is linked to the evaluation of *0-models* was supported by the verification data. *Fewer than n* is often falsely rejected in empty-set situations and takes considerably longer to verify, especially in empty-set situations. Finally, we were able to rule out an alternative pragmatic explanation of these effects according to which processing difficulty was due to a scalar implicature (*fewer than five but not none*). No such pragmatic effect was observed in 11-models for *more than n*.

6 EXPERIMENT 2 – TEASING APART MONOTONICITY AND THE EMPTY SET PROPERTY

The first experiment provided initial evidence that empty-set quantifiers are harder to process than non-empty-set quantifiers. Moreover, it revealed that the evaluation of an empty-set situation in connection with an empty-set quantifier is particularly difficult leading to a substantial increase in error rates relative to situations with positive instances. However, the empty-set quantifier *fewer than five* is of course also monotone decreasing. This raises the question whether it is really the empty-set property that makes a quantifier hard to process, or rather its (downward) monotonicity. In order to further disentangle complexity effects that are due to the presence of empty-set quantifiers from complexity effects due to the monotonicity of the quantifiers, we investigated whether non-monotonic quantifiers which are non-empty-set quantifiers (e.g. *exactly one boy or exactly three boys*) are easier to process than non-monotonic empty-set quantifiers (e.g. *no boy or exactly three boys*). Panels (c) and (d) of Figure 2 further qualify why these quantifiers are an excellent testing ground to tease apart quantificational complexity predicted by the

14 Could the effect observed for 0-models still be a pragmatic effect, albeit of a different nature, namely a presupposition failure? We consider this unlikely for a number of reasons. First of all, 75% acceptance for sentences involving a presupposition failure seems too high given the commonly observed across the board rejection for presupposition failure (e.g., in the case of definite descriptions in contexts lacking a unique referent). Secondly, if this were the case some effect of implicature violation in the 11-models would still be expected given the degraded nature of sentences such as *some elephants have trunks* usually observed in experimental pragmatics (e.g., Bott & Noveck 2004). Thirdly, Exp. 3 shows that the error rates are subject to the overall complexity of the sentences. In the case of *at most one* in doubly quantified sentences error rates drop to chance level. This lends support to an explanation attributing the observed variability in error rates to processing complexity. Another point against a presuppositional analysis of the empty-set effect concerns the fact that the non-emptyness requirement does not seem to project. Considering conditionals with empty-set quantifiers in the antecedent there is the strong intuition that *if fewer than three students came, the course would have been canceled* does not presuppose that some students came. Finally, generalizing this kind of account to all empty-set quantifiers the quantifier *no* should be a lexical item reserved for uses in contexts involving a presupposition failure. This seems very counterintuitive and raises the question why such a lexical item should exist in the first place.

automata account as compared to the present theory. The corresponding automata are almost indistinguishable from each other.

6.1 *Methods*

The present experiment compared the comprehension and verification of non-monotonic empty-set and non-empty-set quantifiers in simply quantified sentences. As in the previous experiment we manipulated two within factors, the factor QUANTIFIER (two levels: *empty-set* vs. *non-empty-set* quantifiers) and the factor MODEL (four levels: *0-models* vs. *1-models* vs. *2-models* vs. *3-models*) in a 2 × 4 factorial design. In line with the results of our first experiment we expected enhanced comprehension and verification difficulty of *empty-set* as compared to *non-empty-set* quantifiers. In particular, in line with Hypothesis 4.1 (*Empty-set effects*) verification difficulty should be observed especially in those trials in which *empty-set* quantifiers have to be evaluated in a *0-model*.

6.1.1 *Participants* 50 participants from Tübingen University took part in the experiment for payment of €5. None of them had participated in the previous experiment. Two participants had more than 20% errors on the filler trials and were excluded from the analysis. The remaining 48 participants (mean age 26.2 years, range 19 – 44 years; 30 female) were all above 80% correct on the fillers. On average they provided correct answers 88.3% of the time. Six participants were randomly assigned to each list.

6.1.2 *Materials* 32 German items were constructed in eight conditions. (19) is a sample item with the *empty set* quantifier *none or three* and the *non-empty set* quantifier *one or three*.

(19) a. Keiner der Punkte oder drei der Punkte sind blau.
No of-the dots or three of-the dots are blue.
'None or three of the dots are blue'.
b. Einer der Punkte oder drei der Punkte sind blau.
One of-the dots or three of-the dots are blue.
'One or three of the dots are blue'.

The sentences always contained the Boolean combination of quantifiers *none or three* and *one or three*. In order to be able to construct minimal pairs which only differ in the sentence initial word, the present experiment employed bare numerals instead of *exactly n*. Previous research has shown that bare numerals are interpreted as *exactly n* (Huang *et al.* (2013), but see, e.g., Marty *et al.* (2013)). It should be the default reading especially in Boolean combinations such as *one or three* where an *at least* reading should be pragmatically blocked (consider the oddness of *at least one or three*). In order to prevent reading time effects due to contradictory subject-verb agreement, the partitive construction *one/none/three of the* was used to ensure that the restrictor noun phrase *dots* was always plural in both disjuncts (for effects of disjuncts of different number in production see, e.g., Haskell *et al.* 2010). In addition, the copula was always plural in order to guarantee local agreement with the immediately preceding disjunct *three of the dots*.

For each sentence pair four pictures were created showing objects in randomized positions. Sample pictures for (19) are provided in Figure 6. Each picture contained three objects in total and all of them were of the relevant category (e.g., dots).

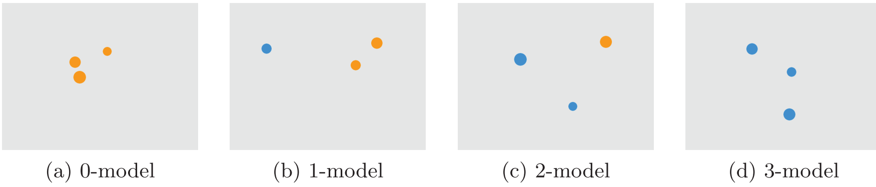


Figure 6 Sample pictures from Experiment 2 for (a) *none* (b) *one or three of the dots are blue*. *0-model* = zero target objects, *1-model* = one target object, *2-model* = two target objects, *3-model* = three target objects.

80 fillers were added with other Boolean combinations of quantifiers such as *three or fewer than two* (40 trials), but also simple quantifiers such as *fewer than three* (40 trials). Half of the fillers were true. In total 50% of the sentences across the experiment required a '*yes, true*' judgment. The total number of objects in the pictures varied in the fillers between two and five. None of the fillers involved an empty-set situation.

A latin square was used to create eight lists such that each participant encountered each item in only one condition while seeing all conditions equally often (as in the previous experiment four data points per condition).

6.1.3 *Procedure* The procedure was the same as in the previous experiment. First participants read a quantified statement which was presented self-paced with word-by-word moving window presentation. After the last word, the sentence disappeared and a new screen with the picture was presented. Participants were instructed to decide as quickly as possible whether the sentence was true or false by pressing the 'yes, true' or 'no, false' button with their left or right index finger. Answers had to be provided within a time limit of four seconds. The response key mapping was again counterbalanced across participants.

The entire experiment was conducted in a single block with individually randomized order of trials. At the beginning of the experiment participants received a short practice of five trials to get used to the task.

6.1.4 *Statistical analysis* Again, statistical analyses were computed for reading times, verification times and error rates. 98.7% of all trials received an answer within the time limit of four seconds. Trials with no answers were treated as missing values in the analysis of judgments and judgment RTs. The reading times and remaining judgment RTs were corrected for outliers by removing all RTs below 100ms and RTs that were more than 2.5 standard deviations above the mean RT for any given participant. This affected 2.5% of the reading time data and 4.1% of the verification data. We computed paired t-tests analyzing the reading times of the nine sentence regions.[15] For the analysis of judgment RTs we statistically analyzed residual RTs taking into account general differences between *yes* and *no* responses (see Clark & Chase 1972; McCloskey & Bigler 1980; Reder, 1982; Singer 1984, a.o.). Residual judgment RTs were fit by linear regression for each participant taking into account the outlier corrected judgment RTs of 'yes, true' and 'no, false' answers for the experimental items in all eight conditions. In the descriptive statistics we will report both raw RTs as well as residual judgment RTs. For the purposes of inferential statistics, residual judgment RTs were analyzed in a repeated measures ANOVA.

15 Because the experimental items were identical except for the nouns and the color adjective again no by-items (t_2) analyses were computed.

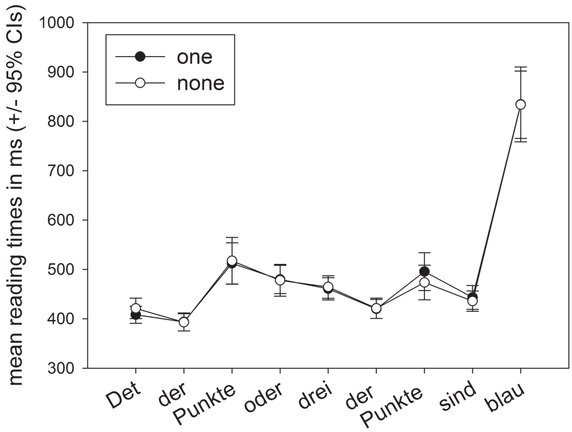


Figure 7 Mean reading times (+95% confidence intervals, computed on the basis of the t_1 analysis) in Experiment 2.

In order to analyze error rates a logit mixed effects model analysis was computed with the fixed effects of QUANTIFIER and MODEL (four levels: *0-model*, *1-model*, *2-model*, *3-model*) and random intercepts and the slopes of QUANTIFIER and MODEL for participants. To break down the interaction we computed separate logit mixed effects models for *0-models* and all other types of models (*1*-, *2*- and *3-models*).

6.2 *Results and Discussion*

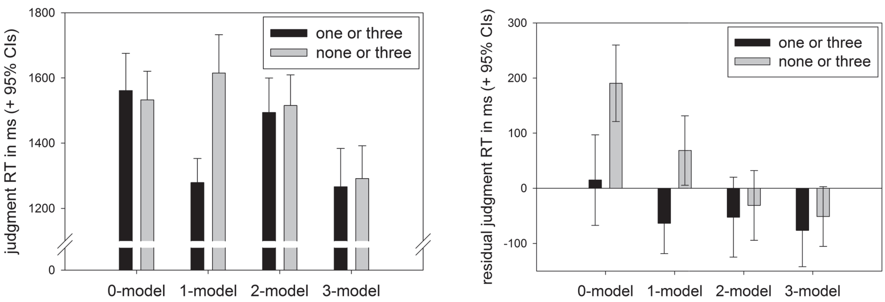
The analysis of reading times did not reveal any reliable differences between *none or three* and *one or three* sentences except for the first word of the sentence (*none* vs. *one*). Figure 7 shows the mean reading times word by word for the entire sentence.

Paired t-tests revealed that the first word *none* was read significantly slower than *one* probably due to lexical differences as, for instance, the number of characters ($t(47) = 2.53; p < .05$), but all subsequent word regions did not show any significant differences ($|t| < 1$; the only exception was region 7 ‘ *dots*’ where we found an effect in the opposite direction ($t(47) = -2.44$)). This is different from what we observed for simply quantified sentences with comparative quantifiers in Experiment 1, where empty set/monotonicity effects were consistently present across both sentence regions.

The observed lack of differences goes against our predictions concerning the comprehension of quantifiers. A possible explanation of this null effect might be that the two quantifiers *none of the dots or three of the dots* and *one of the dots or three of the dots* do not differ in terms of algorithmic complexity while their comparative counterparts do. As described in the introduction empty-set comparative quantifiers require the full set of clauses of the **c-exp** operation. However, the situation may be different for the Aristotelean quantifier *none of the dots*. Here, it is sufficient to check the third clause of the **c-exp** algorithm and return ‘true’ just in case it can be successfully applied and ‘false’ otherwise. In other words, just like *non-empty-set* quantifiers *none* can also be checked with a simplified algorithm using a single rule. What is different from non-empty-set quantifiers is that this algorithm involves clause 3 instead of clause 1 and thus crucially requires the encoding of negative predicate instances as well as polarity reversal. It could therefore well be that the algorithm introduced by *none of the dots* is relatively low in complexity in terms of the number of clauses.

Table 2 Mean error rates in Experiment 2 as a function of QUANTIFIER and MODEL. Correct responses are shown in parentheses.

	none or three	*one or three*
0-model	29.7% (yes)	11.5% (no)
1-model	13.0% (no)	13.5% (yes)
2-model	8.8% (no)	10.4% (no)
3-model	5.7% (yes)	4.7% (yes)



(a) mean judgment RT (b) mean residual judgment RT

Figure 8 Mean judgment times (+/-95% confidence intervals) in Experiment 2. Panel (a) shows the raw judgment RTs, Panel (b) shows the residual judgment RTs after factoring out differences in judgment RT between *yes* and *no* responses.

In contrast to the unexpected findings of the reading time analyses, the analysis of judgment RTs and error rates fully confirmed our predictions. The mean error rates are presented in Table 2. The logit mixed effects model analysis revealed a significant fixed effect for the interaction term QUANTIFIER×MODEL (contrast: *0-model* vs. *all other models*; *estimate* $= -1.66$, $|z| = 2.83$, $p < .01$) which was due to approximately 30% errors in the condition with *none or three* in a *0-model* as compared to less than 15% errors in all other conditions. Pairwise comparisons revealed that the error rate was significantly higher for *none or three* than for *one or three* in the *0-models* (*estimate* $= -1.32$, $|z| = 2.71$, $p < .01$), but no significant difference between the two quantifiers was observed for the other models ($|z| = 1.35$, $p = .18$). Taken together, the analyses of error rates thus provide clear evidence for an empty set effect for *none or three* relative to *one or three*.

Figure 8 presents the mean raw judgment RTs and residual judgment RTs after factoring out the effects of *yes* versus *no* responses. On average, conditions that required a 'yes' answer (Table 2) received a response after 1343 ms and were thus faster than conditions that required a 'no' response, the latter had a mean RT of 1542 ms. The ANOVA of residual judgment RTs revealed a marginally significant interaction between QUANTIFIER and MODEL ($F(3, 141) = 2.58; p = .06$), and significant main effects of QUANTIFIER ($F(3, 141) = 11.63; p < .01$) and MODEL ($F(3, 141) = 7.61; p < .01$). The interaction was due to the fact that *none or three* took longer to verify than *one or three* in *0-models* ($t(47) = 3.80; p < .01$) and in *1-models* ($t(47) = 2.79; p < .01$), whereas there were no significant differences in residual judgment RTs in *2-* and *3-models* (both $t < 1$).

The results of the present experiment provide further evidence for empty-set effects. They further substantiate the claim that the empty-set property makes quantifiers hard to process. In the present experiment both the empty-set quantifier *none or three* and the non-empty-set quantifier *one or three* were non-monotone. Nevertheless similar effects were observed as for *fewer than n* and *more than n* in the previous experiment.[16]

The only difference between the findings of Experiment 1 testing comparative empty-set and non-empty-set quantifiers and those of the present experiment testing Boolean combinations of quantifiers, is the lack of reading time effects in the present experiment. We have already hinted at an explanation but would like to be somewhat more explicit about the potential source of the observed difference. Here is what we consider a plausible explanation: Upon encountering a comparative quantifier of the type *fewer than n* and *more than n* there is a clear difference in algorithmic complexity. Specifying a verification algorithm for *fewer than n* involves the full set of clauses of the **c-exp** expansion operation. This is different for *more than n* which can be evaluated by the simpler **s-exp** operation. This is exactly what we stated in our hypotheses in Section 4. Now, consider the Boolean combinations *none of the A or three of the A are B* and *one of the A or three of the A are B*. The first quantifier may be evaluated as follows. First, check whether exactly three As are B, if yes, return ‘yes’ (the first clause of **s-exp** and **c-exp**), if not, check whether all As are not in B, if yes, return ‘yes’ (the third clause of **c-exp**), otherwise return ‘no’. How about *one or three A are B*? We may define a very similar procedure with the only difference that, this time, we apply simple expansion iteratively: First, check whether three As are B, if yes, return ‘yes’, if not, check whether there is exactly one A that is B, if yes, return ‘yes’, if not, return ‘no’. The two algorithms are arguably of similar complexity as far as only the number of rules is concerned that have to be encoded. Once we consider their execution, however, they differ in complexity since the first algorithm relies on **c-exp** and therefore requires to encode negative predicate instances and to draw a positive conclusion from them. It is plausible that the inherent difference in complexity between the first clause and the third clause only shows up during the actual exection of these clauses, that is during the verification stage. We will come back to this point in the discussion of the next experiment where a similar conclusion will be reached.

7 EXPERIMENT 3 – QUANTIFICATIONAL COMPLEXITY OF DOUBLY QUANTIFIED SENTENCES

Multiply quantified sentences can be evaluated with **s-exp** as long as none of the quantifiers is an empty-set quantifier. In Section 4 two alternative hypotheses were put forward concerning the processing of multiply quantified sentences. According to the first hypothesis (*General difficulty of empty-set quantifiers, Alternative A*) comprehenders switch to the complex expansion algorithm as soon as one quantifier is an empty-set quantifier. The prediction was derived that this would lead to an increase in processing difficulty, but that empty-set effects would not add up. According to the second hypothesis (*General*

16 Some readers may still not be convinced that it is the empty-set property of the quantifiers and not the monotonicity properties of the contituent parts of the Boolean expressions tested in the present experiment. In future work we plan to follow up on this question experimentally by comparing continuous quantifiers such as *between zero and three* and *between one and three*. We are grateful to Stephanie Solt (p.c.) for making this suggestion.

difficulty of empty-set quantifiers, Alternative B), each quantifier in a multiply quantified sentence recursively introduces its own expansion operation, which may be simple or complex. Hence, each empty-set quantifier should contribute difficulty in a cumulative fashion. In order to decide between these two hypotheses the present experiment tested doubly quantified sentences of the form Q_1 *boys tickled* Q_2 *girls*, and manipulated the empty-set property of the two quantifiers. Furthermore, in order to be able to generalize across specific types of quantifiers, it was manipulated whether Q_1 was an Aristotelean quantifier (*jed-* (*each*) vs. *kein-* (*no*)), or a superlative quantifier (*mindestens ein-* (*at least one*) vs. *höchstens ein* (*at most one*)).

Before going into the details of the present experiment an issue orthogonal to our current interests needs to be clarified. Multiply quantified sentences are well known to display scope ambiguities. Accordingly, a sentence such as (20) has a reading in which the quantifiers take linear scope, i.e. there is a particular boy who has tickled all girls, and a reading in which the quantifiers take inverse scope, i.e. every girl has been tickled by a potentially different boy.

(20) Some boy tickled every girl.

Existing psycholinguistic work suggests that inverse scope is more difficult to compute than linear scope (e.g., Anderson 2004). Doubly quantified sentences may therefore have two potentially independent sources of difficulty: a) difficulty induced by the individual quantifiers and b) processing difficulty due to quantifier scope. In order to unambiguously relate processing difficulty to the quantificational complexity of the individual quantifiers, scope has to be held constant across conditions. In German, scope ambiguities are much more constrained than in English. Frey (1993), for instance, observed that German doubly quantified sentences with the canonical subject-before-object word order and without special intonation only allow for a linear interpretation. This observation has been confirmed in psycholinguistic experiments. Bott & Schlotterbeck (2012) compared potentially scope ambiguous doubly quantified sentences with subject-before-object word order, as in (21-a), to unambiguous control conditions in which the quantifiers appeared in clause bounded positions, as in (21-b). In an experiment that probed into which readings are possible during online processing, the two sentence types were indistinguishable from each other. Indication of the inverse scope reading was found only if sentences were presented simultaneously with pictures that disambiguated towards the inverse scope reading and truth-value judgments were given without time limit (what Bott & Schlotterbeck (2012) considered to be 'post-interpretive processing', cf. Caplan & Waters (1999)). We can thus be confident that the doubly quantified sentences in the present study are interpreted with linear scope. Any differences in complexity should therefore be unambiguously linked to the quantificational complexity of the quantifiers tested.

(21) (a) Genau ein Lehrer lobte jeden dieser Schüler.
'Exactly one teacher praised each of these pupils.'
(b) Auf genau einen Lehrer trifft zu, dass er jeden dieser Schüler lobte.
'Exactly one teacher is such that he praised each of these pupils.'

7.1 *Methods*

7.1.1 *Participants* 72 German participants from the University of Tübingen (mean age 25.0 years, range 18 – 43 years, 51 female) took part in the study for payment of €8.

None of them had participated in the previous experiments. An experimental session took approximately 45 minutes. Three participants were randomly assigned to each list.

7.1.2 *Materials* 72 German sentences of the form Q_1 *boys tickled* Q_2 *girls* were constructed. The subject quantifier was either one of the Aristotelian quantificational determiners *jeder* (*each*, non-empty-set) and *kein* (*no*, empty-set) or one of the superlative determiners *mindestens ein* (*at least one*, non-empty-set) and *höchstens ein* (*at most one*, empty-set). The object quantifier was either *mehr als zwei/drei* (*more than two/three*, a non-empty-set quantifier) or *weniger als zwei/drei* (*fewer than two/three*, an empty-set quantifier). (22) is a sample sentence, slashes indicate segmentation for self-paced reading. The underlined noun of the second quantifier is the critical region since this is the earliest point at which a verification algorithm can be fully specified. The first 36 items had *more than two/ fewer than two* as second quantifier, while the remaining 36 had *more than three/ fewer than three*.

(22) Mindestens ein / Junge / kitzelte / mehr / als drei / <u>Mädchen</u>.
At least one / boy / tickled / more / than three / <u>girls</u>.

The comprehension part of the experiment thus involved a factorial 2 (EMPTY-SET Q_1: *empty-set* Q_1 vs. *non-empty-set* Q_1) ×2 (TYPE OF Q_1: *Aristotelian* vs. *superlative*) ×2 (EMPTY-SET Q_2: *empty-set* Q_2 vs. *non-empty-set* Q_2)) within design with eight sentence conditions in total.

For the verification part, each experimental item was paired with three types of set diagrams: 0-, 1- and 3-models which showed no, one or three boys tickling Q_2 girls, respectively. Sample diagrams are presented in Figure 9. Accordingly, the verification part of the experiment employed a factorial 2 (EMPTY-SET Q_1: *empty-set* Q_1 vs. *non-empty-set* Q_1) ×2 (TYPE OF Q_1: *Aristotelian* vs. *superlative*) ×2 (EMPTY-SET Q_2: *empty-set* Q_2 vs. *non-empty-set* Q_2)) ×3 (MODEL: *0-model* vs. *1-model* vs. *3-model*) within design resulting in a total of 24 conditions. Diagrams were constructed in such a way that across sentence conditions always the same set of pictures was used (cf. Figure 9 for an illustration).

The experimental items in the 24 conditions were distributed to 24 lists according to a latin square design. Each list contained each experimental item in only one condition and

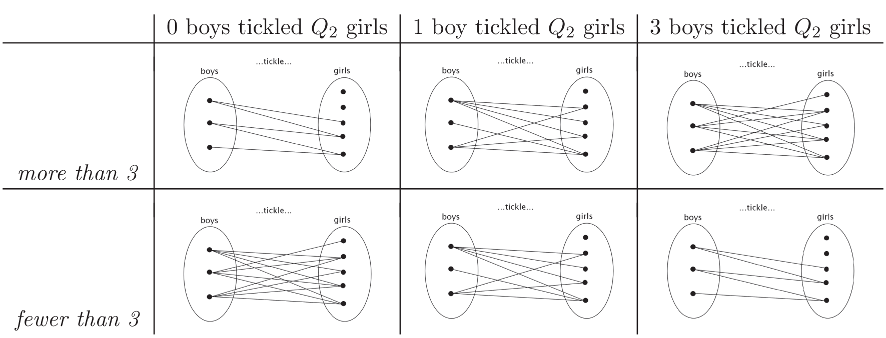


Figure 9 Sample models in Experiment 3 for Q_1 *boys tickled more than/fewer than three girls*. Note: diagrams corresponding to 0- and 3-models were swapped for Q_2= *more than* and Q_2= *fewer than*, but the set of diagrams was kept constant across sentence conditions.

Table 3 Clauses of **c-exp** that had to be applied in Exp. 3. The clauses for **c-exp** of Q_2 are given in parentheses after the clause numbers for **c-exp** of Q_1. Sentence types in boldface indicate quantifier combinations that can be evaluated by **s-exp** in any model. Note: *more* = *more than two/three*, *less* = *fewer than two/three*, *at least* = *at least one*, *at most* = *at most one*; Q_1 = subject quantifier, Q_2 = object quantifier.

	Aristotelean Q_1			
	each...more	each...fewer	no...more	no...fewer
0-model	**s-exp**	4(2/2/2)	3(2/2/2)	3(2/2/2)
1-model	**s-exp**	2(1/2/2)	2(1/2/2)	2(1/2/2)
3-model	**s-exp**	1(1/1/1)	2(1/1/1)	2(1/1/1)
	Superlative Q_1			
	at least...more	at least...fewer	at most...more	at most...fewer
0-model	**s-exp**	4(2/2/2)	3(2/2/2)	3(2/2/2)
1-model	**s-exp**	1(1/2/2)	1(1/2/2)	1(1/2/2)
3-model	**s-exp**	1(1/1/1)	2(1/1/1)	2(1/1/1)

each condition was tested three times per list. 78 additional sentence-picture pairs were included as fillers into each list. Overall 50% of the sentences were true.

7.1.3 *Predictions* We considered which clauses of the**c-exp** verification algorithm had to be applied during the verification of the eight sentence types in each of the three types of models. Care was taken to construct the models in such a way that the applied rules guaranteed maximal comparability between conditions. Table 3 presents a summary of the predicted verification steps in each condition under the assumption of Alternative A of the hypothesis that empty-set quantifiers generally cause difficulty (Hypothesis 4.2). Two sentence types involved combinations of non-empty-set quantifiers, *each* combined with *more than n* and *at least* combined with *more than n*. Since these were the only sentences that could be correctly evaluated by the**s-exp** algorithm we predicted them to be easier to comprehend and verify than all other sentence types. Moreover, Table 3 shows that the steps required for the complex expansion of sentences with an empty-set subject quantifier and a non-empty-set object quantifier were the same as in sentences with two empty-set quantifiers. The complex expansion of the object quantifier only involved the first two clauses: There were always positive predicate instances, thus the expansion of the object quantifier never required the evaluation of the empty set. This allowed us to derive the more specific prediction that the verification of sentences with an empty-set quantifier in subject position should be equally difficult for empty-set and non-empty-set object quantifiers.

Under the assumption of Alternative B (Hypothesis 4.3) each occurrence of a quantifier leads to an application of the **s-exp** or **c-exp** operation. For example, the *each...more* conditions lead to two applications of the **s-exp** operation whereas the *each...fewer* conditions need one application of the **c-exp** operation followed by an application of the **s-exp** operation. Thus, the second alternative (*Alternative B*) predicts combinations of two empty-set quantifiers to be more difficult than combinations of one empty-set and one non-empty-set quantifier.

The type of quantifier manipulation of the first quantifier phrase was included in the experimental design in order to investigate whether the predicted effects of quantificational complexity would interact with other factors known to modulate quantificational complexity. As briefly mentioned above the superlative modifiers *at least* and *at most* have been shown to be inherently more complex than their comparative counterparts *more than* and *fewer than* and probably also than the Aristotelean quantifiers. Geurts & van der Slik (2005) and Geurts *et al.* (2010) showed that these quantifiers turn out to be particularly hard in inference and in verification tasks. Whether this effect is semantic in nature (among others, see Geurts & Nouwen 2007) or whether it originates from the pragmatics of the quantifiers (among others, see Büring 2008; Cummins & Katsos 2010; Schwarz 2013) is not clear, yet. In the present experiment we used TYPE OF FIRST QUANTIFIER simply as a factor inducing quantificational complexity and investigated whether potential effects would take the form of an interaction with the empty-set effects predicted by our theory, or, whether effects of quantificational complexity would add up in a cumulative fashion, i.e. take the form of main effects. Both options seem possible, and we have no specific expectation about the interplay of empty-set effects and superlativity effects.

7.1.4 *Procedure* The procedure was the same as in the previous experiment. Each trial started with a reading stage in which a doubly quantified sentence was read self-paced with moving window presentation. Immediately after the last region had been read, the sentence disappeared and the verification stage started. Participants were shown a set diagram, and they had to provide a truth value judgment. There was no time limit for providing a judgment.

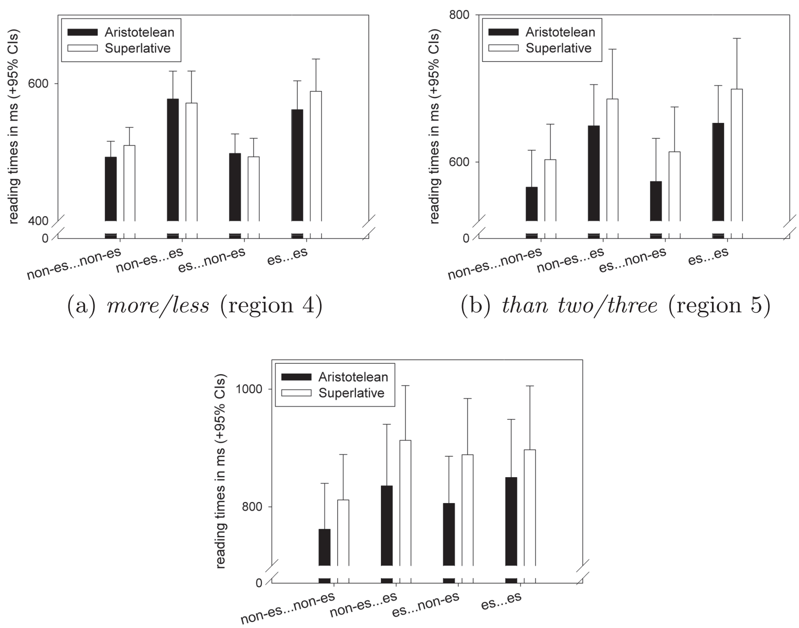
The experiment started with written instructions. Then followed a short practice with eight trials during which participants received feedback. The actual experiment followed in a single block without any feedback. Sentence picture pairs were presented in an individually randomized order.

7.1.5 *Statistical analysis* Again we computed statistical analyses of the reading times, the judgment RTs, and the error rates. Reading times and judgment RTs were corrected for outliers by removing all times below 200ms and judgment RTs more than 2.5 (reading times $\geq$ 3.0) standard deviations above the mean for any given participant and region of interest. This affected 3.3% of the judgment RTs, 0% of the reading time data in the first region, 3.2% in the second region, 3.5% in the third region, 1.7% in the fourth region, 2.9% in the fifth region and 4.1% in the sentence final region. The reading times and judgment RTs were analyzed statistically by computing 2 × 2 × 2 repeated measures ANOVAs with the within factors TYPE OF Q_1, EMPTY-SET Q_1, and EMPTY-SET Q_2, both, by participants (F_1) and by items (F_2). This way, judgment RTs were analyzed for the whole sets of picture materials which - as a whole - were held constant across sentence conditions. Error rates were analyzed by computing two separate logit mixed effects models for Aristotelean Q_1 and for superlative Q_1 with the fixed effects of EMPTY-SET Q_1, EMPTY-SET Q_2 and MODEL (three levels: *0-model*, *1-model*, *3-model*) and random intercepts and the slopes of *model* for participants and items. This was the maximum random effect structure which led to convergence of the computed models.

7.2 *Results and Discussion*

The mean reading times of the critical second quantifier phrase are presented in Figure 10. The first region consisted of the comparative adjective *mehr* or *weniger* (*more*/ *less*). Here the only significant effect was a main effect of EMPTY-SET Q_2 ($F_1(1,71) = 31.56, p < .001$; $F_2(1,71) = 66.19, p < .001$) which was due to the fact that *weniger* was read more slowly than *mehr*. This lexical effect carried over to the following region *als zwei/drei* (*than two/three*) ($F_1(1,71) = 71.57;, p < .001$; $F_2(1,71) = 47.65, p < .001$). Apart from the main effect of EMPTY-SET Q_2, only the main effect of TYPE OF Q_1 turned out to be reliable at this region ($F_1(1,71) = 12.79, p < .001$; $F_2(1,71) = 11.66, p < .01$). Comparative object quantifiers in the context of superlative subject quantifiers were read more slowly than in the context of Aristotelian subject quantifiers. There are a number of potential explanations for this finding. One is that the semantic or pragmatic representations of the sentences involving superlative quantifiers are more complex than that of sentences with Aristotelean quantifiers.

The predicted interaction between EMPTY-SET Q_1 and EMPTY-SET Q_2 (Alternative A, Hypothesis 4.2) was absent on the first two regions of the object quantifier. It did show up, however, on the critical noun region, significantly by participants and marginally significant by items ($F_1(1,71) = 4.32, p < .05$; $F_2(1,71) = 3.01, p = .083$). Note that this is the earliest point at which the verification algorithm could be fully specified. In line with our first hypothesis (Alternative A) that an algorithm based on the complex expansion operation is needed as soon as one of the quantifiers is an empty-set quantifier, conditions with two non-empty-set quantifiers were read faster than all three conditions involving at least one empty-set quantifier. Apart from this interaction only the main effects of EMPTY-SET Q_2 ($F_1(1,71) = 7.62, p < .01$; $F_2(1,71) = 11.33, p < .01$) and of TYPE OF Q_1 were significant at the sentence final region ($F_1(1,71) = 17.62, p < .001$; $F_2(1,71) = 13.25, p < .001$). So, superlative quantifiers seem in fact to be more difficult than Aristotelean quantifiers,



(a) *more/less* (region 4)

(b) *than two/three* (region 5)

(c) *noun* (sentence final region)

Figure 10 Mean reading times (+95% confidence intervals) of the second quantifier regions in Experiment 3. *es* = empty-set quantifier; *non-es* = non-empty-set quantifier.

but this effect did not interact with quantificational complexity induced by the empty set property.

The mean latencies of judgments in the truth-value judgment task are presented in Figure 11. Both empty-set subject and empty-set object quantifiers led to an increase in judgment RTs as reflected by significant main effects of EMPTY-SET Q_1 ($F_1(1,71) = 59.17, p < .001$; $F_2(1,71) = 63.53, p < .001$) and EMPTY-SET Q_2 ($F_1(1,71) = 74.15, p < .001$; $F_2(1,71) = 92.64; p < .001$). However, there was no reliable interaction between the two factors ($F_{1/2} < 3.4$). Note that this is different from the pattern observed in reading times. According to *Alternative A* in Hypothesis 4.2, a single empty-set quantifier already suffices to globally trigger**c-exp**. As can be seen from Table 3, *Alternative A* would let us expect an interaction between *empty set* Q_1 and *empty set* Q_2 because, given that the subject of the test sentence contained an empty-set quantifier, the same clauses were required for empty-set and non-empty-set Q_2. To further test whether this difference was reliable we computed ANOVAs for the subset of conditions with an empty set Q_1. The analysis revealed a highly significant main effect of EMPTY-SET Q_2 ($F_1(1,71) = 42.46, p < .001$; $F_2(1,71) = 24.03, p < .001$) that did not depend on the TYPE OF Q_1 (superaltive vs. Aristotelian), as indicated by a lack of interaction between EMPTY-SET Q_2 and TYPE OF Q_1 ($F_{1/2} < 1$). Thus, the combination of two empty-set quantifiers adds processing difficulty beyond what would be expected on the basis of the first of the two alternative hypotheses.

It thus seems that the second theoretical alternative turns out to be correct. Instead of specifying a **s-exp** or **c-exp** algorithm for the complete sentence, each quantifier seems to translate into a **s-exp** or **c-exp** algorithm of its own. Let us consider the critical two experimental sentence conditions that allow us to decide between the two alternative hypotheses.

Firstly, consider *At most one boy tickled more than three girls*. According to the second alternative (*Alternative B*), verification proceeds as follows. In the first step, the interpreter restricts the domain to boys and girls and evaluates which boys tickle which girls and which do not. The next step is the expansion operation of the object quantifier. Since *more than tree girls* is a non-empty-set quantifier, **s-exp** suffices. To do so, the interpreter has to go through the set of boys and for each of them has to add a plus tuple in case the first clause

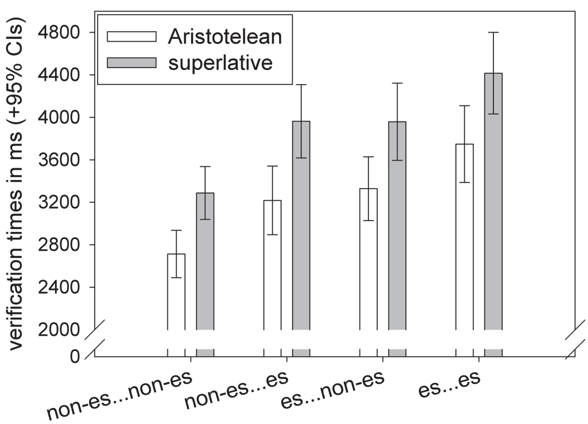


Figure 11 Mean judgment times (+95% confidence intervals, computed on the basis of the F_1 analysis) in Experiment 3. *es* = empty-set quantifier; *non-es* = non-empty-set quantifier.

can applied and a minus tuple, otherwise. Finally, the added tuples have to be evaluated with respect to *at most one boy*, which is an empty-set quantifier, and therefore requires the **c-exp** operation. To summarize then, the sequence of an non-empty-set quantifier and an empty-set quantifier can be evaluated by a sequence of an **s-exp** and a **c-exp** operation, respectively.

This is different for *At most one boy tickled fewer than three girls*. Here we need a sequence of two **c-exp** operations. Such a sequence should be more complex than a sequence of one **s-exp** and a **c-exp** operation. Thus, given theoretical *Alternative B*, it is fully expected that the first sentence type turned out to be easier to verify than the second. By contrast, *Alternative A* would let us expect the two sentence conditions to be equally complex because both require the **c-exp** operation and, moreover, the application of exactly the same clauses of the **c-exp** algorithm.

What is surprising, though, is that the reading times were in line with *Alternative A*, while the verification data provide clear evidence for *Alternative B*. We can only speculate what may have caused this apparent difference between experimental stages. In the discussion of Experiment 2 we already considered the possibility that during comprehension the required rules are put into working memory without specifying a full-blown decision procedure. In the context of the present experiment this would mean that one empty-set quantifier suffices to retrieve the full set of clauses of the **c-exp** operation. The specification of the fully worked out decision procedure would then only happen once the sentence representation can be related to a concrete situation.

Last but not least, consistent with recent proposals from the semantic literature (e.g., Geurts & Nouwen 2007) superlative quantifiers were more difficult to process than Aristotelian quantifiers consistently across the reading and verification stages suggesting that they are inherently more complex. In fact, complexity due to the *type of* Q_1 ($F_1(1,71) = 121.21, p < .001; F_2(1,71) = 110.71, p < .001$), *empty-set* Q_1 ($F_1(1,71) = 59.17, p < .001; F_2(1,71) = 63.53; p < .001$) and *empty-set* Q_2 ($F_1(1,71) = 74.15, p < .001; F_2(1,71) = 92.64, p < .001$) had purely additive effects on judgment RTs (all interactions $F_{1/2} < 3.4$).

The percentages of correct judgments were generally high across all sentence conditions and the three types of models (see Figure 12). An exception to this trend was observed

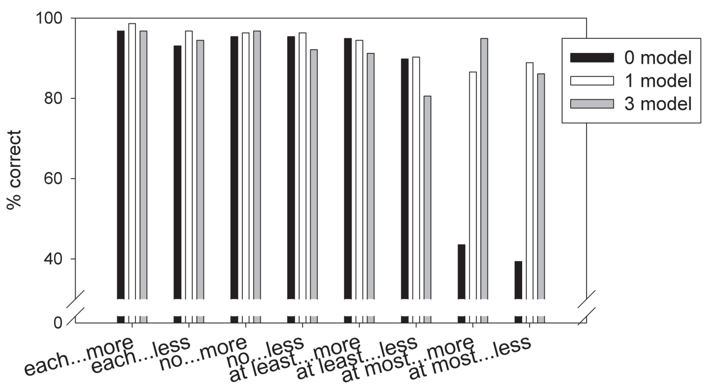


Figure 12 Mean error rates (+95% confidence intervals) in Experiment 3 contingent on sentence type and model. *es* = empty-set Q; *non-es* = non-empty-set Q.

in conditions where *0-models* had to be evaluated in the context of superlative empty-set quantifiers. Only in these conditions, reliable differences in error rates were observed. The logit mixed effects model analysis of the superlative conditions revealed a reliable two-way interaction between *empty-set* Q_1 and *model* (estimate= 3.55, $z = 8.98$, $p < .001$). As in the previous two experiments, proportions of errors were highest in conditions where an empty-set quantifier had to be evaluated in a *0-model*. In these conditions, performance was even significantly below chance level (i.e. 50% acceptance), in particular in sentences with the two empty-set quantifiers *at most* and *fewer than* ($\chi^2(1) = 5.06$; $p < .05$) where only 39.4% of the trials were judged correctly.

Overall, the results of Exp. 3 were consistent with the hypotheses derived from our quantification theory. The observed pattern of reading times supports the distinction between **s-exp** and **c-exp** already during reading and the verification data lend support to the predicted difficulty of empty-set quantifiers when having to evaluate a situation in which the scope of the subject quantifier phrase consists of the empty set. The judgment data suggest that each occurrence of an empty-set quantifier adds to processing difficulty in a cumulative fashion. In the case of superlative quantifiers, processing load becomes so high that interpreters are unable to figure out the meaning of doubly quantified sentences with empty-set quantifiers.[17]

8 GENERAL DISCUSSION

This paper presented a novel theory of quantifier interpretation which predicted that the operation employed in processing empty-set quantifiers is more difficult than that in processing non-empty-set quantifiers. This prediction was confirmed in three experiments that investigated the online comprehension and verification of simply and doubly quantified sentences. All experiments employed a sentence-picture verification task in which participants first read a quantified sentence self-paced and subsequently judged whether the sentence truthfully described a picture.

Experiment 1 established the presence of an empty-set effect in monotone decreasing quantifiers but neither in upward monotone nor non-monotone quantifiers. Furthermore, the experiment provides counter evidence against the alternative explanation that the observed empty-set effect is pragmatic in nature, i.e., that *0-models* were falsely rejected for *fewer than* because *no* would have been a more informative alternative. We did not find a similar effect in *more than* sentences for *11-models* where in fact an *all* sentence would have been more informative.

Experiment 2 disentangled empty-set effects from (downward) monotonicity. We compared ease of interpretation and verification of two non-monotone quantifiers, *none or three* and *one or three*, the first of which is an empty-set quantifiers, whereas the second is not. Again, we observed a clear empty-set effect during verification of *none or three* relative to *one or three* in *0-models*. Surprisingly, however, no empty-set effect showed up during reading.

17 See Bott *et al.* (2013) for evidence that participants are able to deal with *at most* in connection with an empty predicate if the sentences are easy enough (e.g., *at most one dot is blue*) albeit they still make mistakes approximately 25% of the time. Thus, the observed error rates in the range of above 50% in the current experiment indicate that it is by and large cognitive resource limitations which are responsible for the observed pattern of results.

Experiment 3 extended the empirical investigation from simply quantified to doubly quantified sentences. Again, we observed empty-set effects for all empty-set quantifiers included in the experiment, the Aristotelean empty-set quantifier *no*, the comparative empty-set quantifier *fewer than n* and the superlative empty-set quantifier *at most n*. Empty-set effects were present in the interpretation and the verification stage of the experiment. Although there were slight differences between the form these effects took in the reading and in the verification stage, it seems fair to summarize the results of the experiment as indicating that each individual occurrence of an empty-set quantifier independently considerably adds processing difficulty in line with alternative B of the empty-set effect hypothesis.

Taken together, the results of the three experiments strongly support our claim that the empty-set property of quantifiers – among them all DE quantifiers – is a major source of difficulty. In line with the hypotheses summarized in Section 4, we consistently found processing costs related to the empty-set property in general and empty-set situations in particular. An explanation of why this should be the case is unique to the proposed theory of quantification. As discussed on a number of occasions in this article, the semantic automata theory of generalized quantifiers does not predict any of the observed differences between empty-set and non-empty-set quantifiers or monotone decreasing and increasing quantifiers, let alone the observed difficulty in evaluating *0-models*. All upward and downward monotone quantifiers investigated in our study correspond to almost identical acyclic finite state automata (see Figure 2). The parallelism between quantifiers becomes particularly striking when considering the Boolean combinations of quantifiers *none or three* and *one or three* from Experiment 2. Both quantifiers correspond to acyclic finite automata with five states. The only difference is that for *none or three* the initial state is an accepting state, whereas for *one or three* not the first but the second state is an accepting state. Apart from this insignificant difference the automata are exactly the same. Both consist of two accepting and three rejecting states (cf. Panels (c) and (d) of Fig. 2). Within the automata account, it must remain unexplained why the tested quantifiers should differ with respect to their quantificational complexity.

Another important finding that merits rementioning is that our study provides strong evidence for non-monotone (non-empty-set) quantifiers being no more difficult than the tested UE quantifiers. Traditionally, however, these quantifiers receive the most complex truth conditions. For instance, *exactly n* is usually taken to translate into *at least n but not more than n*. The experimental evidence from Exp. 1 and 2 forces us to think about simpler representations for non-monotone quantifiers. It is thus highly welcome that the proposed theory groups non-monotone (non-empty-set) quantifiers together with all the other non-empty-set quantifiers. This sets our proposal clearly apart from other algorithmic approaches such as Beghelli *et al.* (1997).

Our study demonstrates that experimental data can and should play an important role in shaping the way how a theory of natural language quantification must eventually look like in order to be cognitively plausible. The present paper can admittedly only serve as a first step in this direction. We have encountered a few places where the empirical findings call for a revision of the quantification theory as initially proposed. For instance, just making a global distinction between the simple expansion and the complex expansion for entire sentences as *Alternative A* would have it (cf. Bott *et al.* 2013) turned out to be not sufficient because it is incompatible with the cumulative empty-set effects in Experiment 3. Instead, we had to consider the combinations of individual quantifiers in line with Hypothesis 4.3 stated in Section 4. Another empirical finding calling for revisions of the proposed theory were the at first sight incompatible reading time data from Experiment 1 and 2 which revealed a processing difference between upward and downward monotone comparative quantifiers, on the one hand, but a complete lack of effect with respect to empty-set versus non-empty-set Boolean combinations of quantifiers, on the other hand. We suggested that *no* might be a special case because this is the only empty-set quantifier which could also be verified using a simplified verification algorithm compared to the complete complex expansion operation with all four rules. We proposed that for *no* it suffices to only check whether the scope set is empty, hence a single rule is required. We would like to point out once more that the proposed theory still correctly predicts that the required rule, i.e. the third clause of **c-exp**, should be difficult to *verify* because it requires the encoding of negative predicate instances. What we may thus conclude from the reading time data of Exp. 2 is that the interpretation stage may simply consist in updating working memory by adding the required clauses in their order of application. Obviously, further experiments are needed to confirm that this interpretation of the data is on the right track. Finally, Exp. 3 shows that besides the empty-set property other properties such as superlativity clearly influence quantificational complexity. A comprehensive, cognitively realistic quantification theory must therefore explain how the semantic factors interact that bear an influence on quantificational complexity, and, moreover, whether they target the same interpretation steps. Such interpretation steps may include, for instance, the identification of a quantifier's witnesses in contrast to the actual execution of verification algorithms – the topic of the present paper.

9 QUANTIFICATION THEORY – FORMALLY

This section presents the formal details of the theory and proves that it is mathematically well-behaved. We start with the concept of quantifiers in Generalized Quantifier Theory and develop the proposed quantification theory from there.

9.1 *Background*

We propose a model of quantification that applies to exactly those quantifiers which are equivalent to type $\langle 1, 1\rangle$ quantifiers in *Generalized Quantifier Theory* (GQT) that are closed under isomorphism (a fundamental and characteristic property of quantification often abbreviated as Isom), conservative (Conserv) and domain independent (referred to as Ext for extension, cf. Peters & Westerståhl (2006) for a comprehensive overview of GQT). For this reason, the proposal gives an explanation of the observation that quantifiers in natural language generally have these properties. In particular, as is shown below, the domain-relativizing function of the restrictor argument is explained.

In GQT, quantifiers are second order relations over first order relations over individuals (definition adopted from Peters & Westerståhl 2006, p. 64).

Definition 9.1 (**generalized quantifiers satisfying** isom). Let $\tau = \langle n_1, \dots, n_k\rangle$ be a sequence of positive, non-zero integers. A generalized quantifier Q of type τ associates with each set M a k-ary second order relation Q_M over M, where the i-th argument of Q_M is a n_i-ary relation over individuals in M such that the following holds true: If the structure $\mathcal{M} = (M, R_1, \dots R_k)$, with $R_i \subseteq M^{n_i}$, is isomorphic to another structure $\mathcal{M}' = (M', R'_1, \dots R'_k)$, then

$$Q_M(R_1, \dots R_k) \Longleftrightarrow Q_{M'}(R'_1, \dots R'_k).$$

This concept of generalized quantification can also be used to build formulas of an enriched predicate logic. To illustrate, assume that Q is a generalized quantifier of type $\langle 1,1\rangle$. Using the symbol Q as a variable binding operator we can build quantified formulas of the form $\mathrm{Qx,y}(\phi_1(\mathrm{x}, \mathrm{a}_1, \ldots, \mathrm{a}_{\mathrm{m}_1}), \phi_2(\mathrm{y}, \mathrm{b}_1, \ldots, \mathrm{b}_{\mathrm{m}_2}))$. Now let $\phi(\mathrm{z}, \mathrm{c}_1, \ldots, \mathrm{c}_\mathrm{m})^{\mathcal{M},\mathrm{z}}$ denote the set $\{c_0 \in M : \mathcal{M} \models \phi(\mathrm{c}_0, \mathrm{c}_1, \ldots, \mathrm{c}_\mathrm{m})\}$, where the constant symbol c_0 is interpreted as constant c_0 in $\mathcal{M}$. Then, formulas like the above are interpreted in model $\mathcal{M}$ using the rule:

$$\mathcal{M} \models Qx,y\big(\phi_1(x, a_1, \ldots, a_{m_1}), \phi_2(y, b_1, \ldots, b_{m_2})\big) \Leftrightarrow$$
$$Q_M\Big(\phi_1(x, a_1, \ldots, a_{m_1})^{\mathcal{M},x}, \phi_2(y, b_1, \ldots, b_{m_2})^{\mathcal{M},y}\Big)$$

The properties Ext and Conserv are defined as follows (adopted from: Peters & Westerståhl 2006, p. 105 and p. 138, respectively)

Definition 9.2 (ext **and** conserv **of type** $\langle 1,1\rangle$ **quantifiers**). A type $\langle 1,1\rangle$ quantifier Q satisfies *extension* (Ext) iff the following holds:

$$\textit{If } A, B \subseteq M, \textit{and } M \subseteq M', \textit{then } Q_M(A,B) \Leftrightarrow Q_{M'}(A,B).$$

A type $\langle 1,1\rangle$ quantifier Q is called *conservative* (Conserv) iff, for all M, and all $A, B \subseteq M$,

$$Q_M(A,B) \Longleftrightarrow Q_M(A, A \cap B).$$

Using the GQT framework, natural language determiners are standardly assumed to denote type $\langle 1,1\rangle$ quantifiers, i.e. binary functions from subsets of the domain of individuals M into the set of truth values, satisfying conserv and ext. The first argument of such quantifiers plays a special role: We can think of them as relativizations of type $\langle 1\rangle$ quantifiers, where the extra argument restricts the universe.

Definition 9.3 (**relativization of type** $\langle 1\rangle$ **quantifiers**). Let Q be a generalized quantifier of type $\langle 1\rangle$. Then, Q^{rel} is of type $\langle 1,1\rangle$ and is defined as follows, for $A, B \subseteq M$:

$$(Q^{rel})_M(A,B) \Longleftrightarrow Q_A(A \cap B).$$

This relativization operation is central to understanding the domain relativizing function of the restrictor argument of natural language determiners. Peters & Westerståhl (2006, p. 141f.) prove the following proposition regarding this operation.

Proposition 9.4. The operation $\cdot^{rel}$ is a bijection between the class of type $\langle 1\rangle$ quantifiers and the class of type $\langle 1,1\rangle$ quantifiers that satisfy conserv and ext.

9.2 *The present approach*

In what follows we shall break with GQT tradition in fundamental respects and think of determiners as denoting certain types of unary, instead of binary functions. These functions, which we call *w-functions* (short for witness set functions), map a restrictor set to a tuple consisting of the restrictor itself and a set of *witness sets*:

Definition 9.5 (**w-function, w-quantifier, witness sets**). Let M be any set (called the domain of entities or universe) and $\mathcal{P}(M)$ denote its power set. A function

$$q : \mathcal{P}(M) \to \mathcal{P}(M) \times \mathcal{P}(\mathcal{P}(M))$$

is called a w-function if for any $A \subseteq M$ there is a $W \subseteq \mathcal{P}(A)$, such that $q(A) = \langle A, W \rangle$. If q is a w-function and $A \subseteq M$, then $q(A) = \langle A, W \rangle$ is called a w-quantifier, and W the set of witness sets of q at A.

Example 9.6 (w-quantifiers).

$$
\begin{aligned}
[\![\text{some}]\!](A) &= \langle A, \{X : X \subseteq A \wedge |X| \geq 1\}\rangle \\
[\![\text{every}]\!](A) &= \langle A, \{X : X \subseteq A \wedge X = A\}\rangle \\
[\![\text{not every}]\!](A) &= \langle A, \{X : X \subseteq A \wedge X \neq A\}\rangle \\
[\![\text{no}]\!](A) &= \langle A, \{X : X \subseteq A \wedge |X| = 0\}\rangle \\
[\![\text{more than two}]\!](A) &= \langle A, \{X : X \subseteq A \wedge |X| > 2\}\rangle \\
[\![\text{less than two}]\!](A) &= \langle A, \{X : X \subseteq A \wedge |X| < 2\}\rangle \\
[\![\text{most}]\!](A) &= \langle A, \{X : X \subseteq A \wedge |A \cap X| > |A - X|\}\rangle \\
[\![\text{no or exactly three}]\!](A) &= \langle A, \{X : X \subseteq A \wedge (|X| = 0 \vee |X| = 3)\}\rangle
\end{aligned}
$$

Although our proposal departs from GQT tradition, w-quantifiers can be translated into generalized quantifiers of type $\langle 1, 1\rangle$ that satisfy Conserv and Ext and *vice versa*. This is stated more formally in the following proposition.

Proposition 9.7. There is a one-to-one correspondence between the class of w-functions (as defined in 9.5) and the the class of type $\langle 1, 1\rangle$ generalized quantifiers that satisfy Conserv and Ext.

Proof. Given the above Proposition 9.4, it suffices to show that there is a bijection between the type $\langle 1\rangle$ quantifiers and the w-functions. It is easy to see that such a bijection is provided by the mapping that maps each type $\langle 1\rangle$ quantifier Q to a w-function

$$q : A \mapsto \big\langle A, \{X : Q_A(X)\}\big\rangle.$$

□

Next, we define Boolean combinations of w-quantifiers. Given a w-quantifier $q(A) = \langle A, W\rangle$, the negation $\neg q(A)$ is defined as that pair consisting of the same restrictor set A and the sets of those subsets of A which are not in W, i.e. $\neg q(A) = \langle A, \mathcal{P}(A) - W\rangle = \langle A, \overline{W}\rangle$ for short. Here are two examples:

$$
\begin{aligned}
\neg[\![\textit{every}]\!](A) &= \langle A, \{X : X \subseteq A \wedge X \neq A\}\rangle \\
\neg[\![\textit{fewer than five}]\!](A) &= \langle A, \{X : X \subseteq A \wedge |X| \geq 5\}\rangle.
\end{aligned}
$$

The conjunction of two DP-denotations $\langle A, W\rangle$ and $\langle A', W'\rangle$ is the pair consisting of the union $A \cup A'$ of the two restrictor sets, and the set of pairwise unions of witness sets in W and W', i.e. $\wedge(\langle A, W\rangle, \langle A', W'\rangle) = \langle A \cup A', \{w \cup w' : w \in W \wedge w' \in W'\}\rangle$. For example, the denotation of *every referee and at most one paper* in a model with $[\![referee]\!] = \{r_1, r_2, r_3\}$ and $[\![paper]\!] = \{p_1, p_2, p_3\}$ is:

$$\begin{aligned}\wedge([\![every\ referee]\!], [\![at\ most\ one\ paper]\!]) = \Big\langle &\{r_1, r_2, r_3, p_1, p_2, p_3\},\\ &\big\{\{r_1, r_2, r_3\}, \{r_1, r_2, r_3, p_1\},\\ &\{r_1, r_2, r_3, p_2\}, \{r_1, r_2, r_3, p_3\}\big\}\Big\rangle.\end{aligned}$$

The (right) monotonicity properties of type $\langle 1, 1\rangle$ quantifiers correspond to the following monotonicity properties of w-quantifiers:

Definition 9.8 (**increasing, decreasing and non-monotone w-quantifiers**). A w-quantifier $q(A) = \langle A, W\rangle$ is called monotone increasing iff for every $X, X' \subseteq A$ we have $X \in W \wedge X \subseteq X' \rightarrow X' \in W$. A w-quantifier $q(A) = \langle A, W\rangle$ is called monotone decreasing iff $X \in W \wedge X' \subseteq X \rightarrow X' \in W$. A w-quantifier is non-monotone if it is neither monotone increasing nor monotone decreasing.

The empty-set property of quantifiers is closely related to monotonicity, albeit the two properties are not identical.

Definition 9.9 ((**non-) empty-set quantifiers**). Let M be the domain of entities, and q a w-function: Then q is an empty-set w-function iff for some non-empty subset A of M $q(A) = \langle A, W\rangle$ and W contains the empty set. Otherwise, we call q a non-empty-set w-function. If q is a (non-)empty-set w-function, $q(A)$ is called a (non-)empty-set quantifier.

Note that according to this definition a determiner like *every* does not denote an empty-set w-function. This is true despite the fact that in the case of an empty restrictor set, the quantifier is trivially true[18] . *Every* does not denote an empty-set w-function because for all non-empty restrictor sets A the empty set is not included in the set of witness sets. By contrast, a determiner such as *all but at most n* denotes an empty-set w-function, even though it is upward monotone. This is because for restrictor sets of cardinalities of up to n, *all but at most n A are B* is true if none of the As is in B. For instance, *all but at most five dots are blue* is true in a situation with five dots, i.e. the restrictor set is non-empty, and all of them are of a different colour than blue. Again, these cases may be odd due to a precondition that the restrictor set must consist of at least $n + 1$ elements. Exceptive expressions have received considerable discussion in the literature, which is well beyond the scope of the present paper Hoeksema (1990); von Fintel (1993); Moltmann (1995); García-Álvarez (2008).

All monotone decreasing quantifiers are empty-set quantifiers. The class of monotone increasing and non-monotone quantifiers, however, includes both empty-set quantifiers as well as non-empty-set quantifiers. For instance, *exactly one or exactly three* is a non empty-set quantifier, but *none or exactly three referees* is an empty-set quantifier. The situation is less clear for monotone increasing quantifiers but it is at least not inconceivable that *all but*

18 Here we ignore the requirement that the restrictor must not be empty. This we take to be a presupposition of NL quantifiers, following Geurts & van der Sandt (1999).

at most hundred papers is an empty-set quantifier, while *more than five papers* or *all papers* are non-empty-set quantifiers.

9.3 *Expanding predicates with quantifiers*

One main strength of the proposed approach to quantification is that it also sheds new light on the processing costs of quantifiers and quantified sentences. As proposition 9.28 below shows, if a quantifier does not contain the empty set as a witness set, then the simple expansion operation sketched in Section 2, i.e. the operation which adds only positive information based on positive information, is sufficient to expand any predicate by this quantifier. For empty-set quantifiers, on the other hand, we need a more complex expansion operation. We now introduce the necessary background for our main proposition 9.28, starting with our conception of predicates.

We think of n-ary predicates as sets of n-tuples consisting of elements of the universe M or of w-quantifiers. We allow predicates to apply to both individuals and w-quantifiers alike. This will become relevant in the definitions of the two expansion operations below.

Definition 9.10 (**n-ary predicate, simple n-ary predicate**). A subset P of $(M \cup (\mathcal{P}(M) \times \mathcal{P}(\mathcal{P}(M))))^n$, with $n \geq 1$, is called an n-ary predicate. An n-ary predicate not containing any w-quantifiers, i.e. a subset P of M^n, is called a simple n-ary predicate.

Example 9.11. Let $M = \{a, b, c, d, e, f\}$ and $A = \{a, b, c\}$. P_1 and P_2 are examples of unary predicates, and P_3, P_4 and P_5 are examples of binary predicates. While P_1 and P_3 are simple unary and binary predicates, the others are not:

$$P_1 = \{\langle a\rangle, \langle d\rangle, \langle f\rangle\}$$

$$P_2 = \{\langle [\![\textit{exactly three}]\!](A)\rangle, \langle e\rangle\}$$

$$P_3 = \{\langle a, e\rangle, \langle f, b\rangle, \langle c, d\rangle\}$$

$$P_4 = \{\langle [\![\textit{every}]\!](A), c\rangle, \langle b, [\![\textit{some}]\!](A)\rangle\}$$

$$P_5 = \{\langle [\![\textit{every}]\!](A), [\![\textit{some}]\!](A)\rangle\}$$

Based on this notion of an n-ary predicate, we define the notion of a polarity relation. Specifically, any simple n-ary predicate P is associated with a polarity relation P^*, which essentially is the characteristic function of P. This allows for the explicit expression of positive and negative predicate instances in the extension of predicates. In order to keep the definition of polarity relations concise, we use the following notation.

Notation 9.12: blackLet σ be an n-tuple.

- $\pi_i(\sigma)$ denotes the i-th element of σ,
- σ^+ stands for the tuple $\langle \pi_1(\sigma), \ldots, \pi_n(\sigma), + \rangle$ and
- σ^- stands for the tuple $\langle \pi_1(\sigma), \ldots, \pi_n(\sigma), - \rangle$.

Definition 9.13 (**n-ary polarity relation, polarity relation of a simple n-ary predicate**). Let A and B be disjoint subsets of $(M \cup (\mathcal{P}(M) \times \mathcal{P}(\mathcal{P}(M)))^n$. Then:

$$R = \{\sigma^+ : \sigma \in A\} \cup \{\sigma^- : \sigma \in B\}$$

is called a polarity relation. Moreover, any n-ary simple predicate $P \subseteq M^n$ is associated with a polarity relation

$$P^* = \{\sigma^+ : \sigma \in P\} \cup \{\sigma^- : \sigma \in M^n \wedge \sigma \notin P\}.$$

Example 9.14 (polarity relation). Let $M = \{a, b\}$, and let $P = \{\langle a, a\rangle, \langle b, b\rangle\}$. Then the polarity relation P^* of P is $P^* = \{\langle a, a, +\rangle, \langle a, b, -\rangle, \langle b, a, -\rangle, \langle b, b, +\rangle\}$.

Next, we define the set of i-fillers $[\sigma]_i^R$ of a tuple $\sigma \in R$ as the set of individuals in the i-th position of any i-variant $\tau \in R$ of σ (τ is an i-variant of σ iff τ and σ differ at most at the i-th position). The set of i-fillers of a tuple at a given position provides us with the set of individuals which do or do not participate in the relation R. Furthermore, when we use i-fillers below to expand polarity relations with w-quantifiers, the set of i-fillers always contains all the elements of the restrictor set A_i that participate or do not participate in the relation R. This way maximality is guaranteed, an important condition for any semantic account of quantification in terms of witness sets Beghelli *et al.* (1997); Szabolcsi (2010); Robaldo *et al.* (2014).

Definition 9.15 (***i*-fillers**). Let R be an n-ary polarity relation and let $\sigma \sim_i \tau$ hold iff σ and τ differ at most at the i-th element. For every $\sigma \in R$ and any integer $1 \leq i \leq n$, we call $[\sigma]_i^R = \{\pi_i(\tau) : \tau \in R \wedge \tau \sim_i \sigma\}$ the i-fillers of σ in P.

Example 9.16 (i-fillers). Let $R = \{\langle a, a, +\rangle, \langle a, b, +\rangle, \langle b, a, -\rangle, \langle b, b, +\rangle\}$. Then:

$$[\langle a, a, +\rangle]_1^R = \{a\}$$
$$[\langle a, a, +\rangle]_2^R = \{a, b\}$$
$$[\langle b, a, -\rangle]_1^R = \{b\}$$
$$[\langle b, a, -\rangle]_2^R = \{a\}$$

As noted above the way the set of i-fillers is defined guarantees that we are always dealing with the maximal set of individuals that form positive instances of a predicate. Maximality is an important condition for any quantifier theory grounded in witness sets. For instance, Szabolcsi (2010, p. 56) points out that the contribution of monotone increasing quantifiers can be formulated in the following form of existential quantification over witness sets:

(23) (a) *At least two men walk* = There is a set of men with cardinality at least two such that all its elements walk.

(b) *At most two men walk* ≠ There is a set of men with cardinality at most two such that all its elements walk.

(c) *Exactly two men walk* ≠ There is a set of men with cardinality exactly two such that all its elements walk.

Szabolcsi (2010, 57) concludes that "sentences involving non-increasing quantifiers can only be rephrased using existential quantification over sets of a given size if a maximality condition is added, i.e. if we guarantee that we are inspecting the largest possible situation." This problem is significant, Szabolcsi claims, because "existential quantification over sets is often seen as a desirable tool for formalizing certain meanings." On the present account, maximality is built into the semantic system right from the start. Sentence (23-b), for instance, in a situation with three walking men leads to the construction of the set of *i*-fillers consisting of three men. No smaller set is allowed because the construction of this set according to Definition 9.15 does not involve existential quantification, but all individuals with the relevant property are considered.

The set of *i*-fillers is used to expand a predicate with a w-quantifier. We therefore define the expansion operation – the key notion regarding quantificational complexity – which either adds a positively marked tuple including the w-quantifier in case quantifier evaluation is successful or a negative tuple if evaluation fails. We restrict the expansion operation to cases with no w-quantifiers contained in the respective argument position of the polarity relation because we want to ensure that the expansion operation never produces inconsistencies.

In line with the results of Exp. 3 reported in Section 7, we deviate from the exposition in Section 2 and define polarity relations (and the simple expansion operation, Definition 9.18) in such a way that they are consistent with Alternative B of the *Difficulty-of-Empty-Quantifiers Hypothesis* from Section 4. Note that the recursive construction of complex expansion algorithms makes it necessary to encode negative information when processing a non-empty-set quantifier in order to pass this information on to an empty-set quantifier with wider scope. For Alternative A of this hypothesis, the first of the three clauses below would suffice and we could also expand a predicate directly instead of resorting to its polarity relation. This is how we presented the the **s-exp** operation in Section 2.

Definition 9.17 (**Substitution**). Let R be an n-ary polarity relation. Then, for any $\sigma \in R$ and any $1 \leq i \leq n$, $\sigma[i/x]$ is the result of replacing the i-th element of σ by x.

Definition 9.18 (**simple *i*-expansion, successful simple *i*-expansion**). Let q be a non-empty-set w-function, R be a non-empty polarity relation and $A \subseteq M$. If the set $\{\pi_i(\tau) : \tau \in R\}$ is a subset of M (i.e. no tuple in the polarity relation contains a w-quantifier at position i), then the simple expansion of R by $q(A) = \langle A, W \rangle$ at position i, written as $\mathbf{s\text{-}exp}_i(q(A), R)$, is the smallest set Q such that $R \subseteq Q$ and the following clauses hold:

1 $\sigma^+ \in R \wedge ([\sigma^+]_i^R \cap A) \in W \rightarrow \sigma^+[i/q(A)] \in Q$

1 $\sigma^+ \in R \wedge ([\sigma^+]_i^R \cap A) \notin W \rightarrow \sigma^-[i/q(A)] \in Q$

3 $\sigma^- \in R \wedge ([\sigma^-]_i^R \cap A) = A \rightarrow \sigma^-[i/q(A)] \in Q$

The simple expansion of R by $q(A)$ at position i is called successful iff there is a σ^+ in R such that $\sigma^+[i/q(A)] \in \mathbf{s\text{-}exp}_i(q(A), R)$. Otherwise, we say the expansion fails.

Example 9.19. Consider the sentence:

(24) More than two squares are pink

in the following model: $M = \{a, b, c, d, e\}$, $[\![square]\!] = \{a, b, c, d\}$, $[\![pink]\!] = \{a, b, c\}$. The polarity relation of $[\![pink]\!]$ is $[\![pink]\!]^* = \{\langle a, + \rangle, \langle b, + \rangle, \langle c, + \rangle, \langle d, - \rangle, \langle e, - \rangle\}$, the quantifier $\exists^{>2}([\![square]\!]) = \langle [\![square]\!], \{\{a, b, c\}, \{a, b, d\}, \{a, c, d\}, \{b, c, d\}, \{a, b, c, d\}\} \rangle$. Note that the empty set is not a witness set of $\exists^{>2}([\![square]\!])$, and therefore the simple expansion of $[\![pink]\!]^*$ by $\exists^{>2}([\![square]\!])$ at position 1 is:

$$\mathbf{s\text{-}exp}_1(\exists^{>2}([\![square]\!]), [\![pink]\!]^*) = [\![pink]\!]^* \cup \{\langle \exists^{>2}([\![square]\!]), + \rangle\}$$

since $\langle a,+\rangle_1^{[\![pink]\!]^*} \cap [\![square]\!] = \{a,b,c\}$ and $\{a,b,c\}$ is a witness set of $\exists^{>2}([\![square]\!])$ (the antecedent of clause 1 is true).

If, on the other hand, the w-function q is an empty-set function, then simple expansion may yield the wrong result. This is why we restrict it to non-empty-set quantifiers. To see why, consider the following example.

Example 9.20. The sentence

(25) Less than two squares are pink.

is true in the model $M = \{a,b,c,d,e\}$, $[\![square]\!] = \{a,b,c,d\}$, $[\![pink]\!] = \{\}$. The polarity relation $[\![pink]\!]^* = \{\langle a,-\rangle, \langle b,-\rangle, \langle c,-\rangle, \langle d,-\rangle, \langle e,-\rangle\}$ contains only negative tuples, and therefore the first clause does not apply, and consequently $\langle \exists^{<2}([\![square]\!]),+\rangle$ cannot be added. The only clause that applies is clause 3, and therefore simple expansion incorrectly adds the pair $\langle \exists^{<2}([\![square]\!]),-\rangle$.

This restriction of the simple expansion operation motivates the introduction of the complex expansion operation.

Definition 9.21 (**complex *i*-expansion**). Let q be a w-function, R be a polarity relation and $A \subseteq M$. If the set $\{\pi_i(\tau) : \tau \in R\}$ is a subset of M, then the complex expansion of R by $q(A) = \langle A, W\rangle$ at position i, written as $\mathbf{c\text{-}exp}_i(q(A),R)$ is the smallest set Q such that $R \subseteq Q$ and clauses 1 through 4 hold:

1 $\sigma^+ \in R \wedge ([\sigma^+]_i^R \cap A) \in W \rightarrow \sigma^+[i/q(A)] \in Q$
1 $\sigma^+ \in R \wedge ([\sigma^+]_i^R \cap A) \notin W \rightarrow \sigma^-[i/q(A)] \in Q$
3 $\sigma^- \in R \wedge ([\sigma^-]_i^R \cap A) = A \wedge \emptyset \in W \rightarrow \sigma^+[i/q(A)] \in Q$
4 $\sigma^- \in R \wedge ([\sigma^-]_i^R \cap A) = A \wedge \emptyset \notin W \rightarrow \sigma^-[i/q(A)] \in Q$

Example 9.22. Returning to the previous example, we can now show that complex expansion yields the correct result, namely the addition of the pair $\langle \exists^{<2}([\![square]\!]),+\rangle$.

(26) Less than two squares are pink.

First, note that the set $[\langle a,-\rangle]_1^{[\![pink]\!]^*} \cap [\![square]\!]$ of squares which are 1-fillers of the tuple $\langle a,-\rangle$ is $\{a,b,c,d\}$ which is precisely the extension of the restrictor $[\![square]\!]$. Secondly, note that the empty set is a witness set of the w-quantifier $\exists^{<2}([\![square]\!]) = \langle [\![square]\!], \{\emptyset,\{a\},\{b\},\{c\},\{d\}\}\rangle$. Therefore the antecedent of clause 3 of complex expansion is true, and consequently $\langle \exists^{<2}([\![square]\!]),+\rangle \in \mathbf{c\text{-}exp}_1(\exists^{<2}([\![square]\!]),[\![pink]\!]^*)$. Note finally, that given the polarity relation $[\![pink]\!]^*$, the antecendents of clauses 1, 2 and 4 cannot apply, and therefore no other tuple can be added.[19]

The following proposition shows that complex expansion is equivalent to simple expansion, as long as a given w-quantifier does not contain the empty set as a witness set.

19 Further examples illustrating the **c-exp** operation for doubly quantified sentences have been discussed in Section 2. Also, cf. the informal characterization of the **c-exp** operation in (7), ibid.

Proposition 9.23. Let $A \subseteq M$ be a non-empty subset of the domain, $q(A)$ be a non-empty-set w-quantifier, and let R be an n-ary polarity relation. Then for any positive, non-zero integer $i \leq n$:

$$\textbf{s-exp}_i(q(A), R) = \textbf{c-exp}_i(q(A), R)$$

Proof. Since the first two clauses of simple and complex expansion are the same, it follows that **for any positive tuple** $\sigma^+ \in R$ **we have**:

$$\sigma^+[i/q(A)] \in \textbf{s-exp}_i(q(A), R) \Leftrightarrow \sigma^+[i/q(A)] \in \textbf{c-exp}_i(q(A), R)$$

due to clause 1, and

$$\sigma^-[i/q(A)] \in \textbf{s-exp}_i(q(A), R) \Leftrightarrow \sigma^-[i/q(A)] \in \textbf{c-exp}_i(q(A), R)$$

due to clause 2. Next, note that the antecedent of clause 3 of complex expansion can never be true, because for every non-empty $A \subseteq M$ the empty set is never a witness set of the w-quantifier $q(A)$. For the same reason the conditions and the effect of clause 3 of complex expansion and clause 4 of simple expansion are the same. Therefore, **for any negative tuple** $\sigma^- \in R$ we have that:

$$\sigma^-[i/q(A)] \in \textbf{s-exp}_i(q(A), R) \Leftrightarrow \sigma^-[i/q(A)] \in \textbf{c-exp}_i(q(A), R)$$

□

The complex i-expansion of polarity relation R by a quantifier $q(A)$ at position i is consistent: It adds $\sigma^+[i/q(A)]$ to P^* just in case it does not add $\sigma^-[i/q(A)]$.

Proposition 9.24. Let R be an n-ary polarity relation, $\{\pi_i(\tau) : \tau \in R\} \subseteq M$ and $q(A)$ be any w-quantifier. Then, for any $\sigma^+ \in R$ and for any $\sigma^- \in R$ we have:

$$\sigma^+[i/q(A)] \in \textbf{c-exp}_i\,(q(A), R) \Leftrightarrow \sigma^-[i/q(A)] \notin \textbf{c-exp}_i\,(q(A), R)\,.$$

Proof. $\Rightarrow$: We assume that $\sigma^+[i/q(A)] \in \textbf{c-exp}_i(q(A), R)$. First, we consider the case that $\sigma^+ \in R$. By the definition of polarity relations it follows from $\sigma^+ \in R$ that $\sigma^- \notin R$. Therefore, clauses 3 and 4 cannot apply. Since $\sigma^+ \in R$, $\sigma^+[i/q(A)] \notin R$ and $\sigma^+[i/q(A)] \in \textbf{c-exp}_i(q(A), R)$, we can conclude that clause 1 must have applied, and therefore clause 2 cannot have applied. Since clauses 2 and 4 cannot have applied, it follows that $\sigma^-[i/q(A)] \notin \textbf{c-exp}_i(q(A), R)$. For the case that $\sigma^- \in R$, we proceed analogously.
$\Leftarrow$: We assume that $\sigma^-[i/q(A)] \notin \textbf{c-exp}_i(q(A), R)$ and consider the case that $\sigma^- \in R$ first. From $\sigma^- \in R$ it follows that clause 1 and 2 cannot apply. Since $\sigma^-[i/q(A)] \notin \textbf{c-exp}_i(q(A), R)$, it follows that clause 4 cannot have applied either. Hence, clause 3 applies. Therefore, $\sigma^+[i/q(A)] \in \textbf{c-exp}_i(q(A), R)$. For the case that $\sigma^+ \in R$, we proceed analogously.

□

So far, we have dealt with the expansion of polarity relations by a single w-quantifier at some position i. The next two definitions allow us to deal with iterations of quantifiers.

Definition 9.25 (**semantic role assignment store**). A semantic role assignment store S of an n-ary verb (phrase) is a set of pairs $\langle i, a\rangle$, where

(1) a is a w-quantifier or an element of the domain M,
(2) $i > 0$ is a natural number expressing the semantic role assigned to a, and
(3) for every $s, s' \in S$, $\pi_1(s) \neq \pi_1(s')$, i.e. no two arguments have the same semantic role.

Definition 9.26 (**syntax and semantics**). Let d be an expression denoting some w-quantifier $q(A)$, let R be an n-ary polarity relation ($n \geq 1$), let S be a store, and let v be an expression denoting the pair $\langle S, R\rangle$. Then for every $1 \leq i \leq n$, $d \bullet_i v$ is defined iff for every element $s \in S$, $\pi_1(s) \neq i$, i.e. no element of S has the semantic role i. If $d \bullet_i v$ is defined, then it denotes the pair $\langle S \cup \{\langle i, q(A)\rangle\}, R\rangle$, and i is called the semantic role of $q(A)$.

In other words, the combination of a (determiner) phrase denoting a w-quantifier $q(A)$ and a verb (phrase) denoting the pair $\langle S, R\rangle$ results in indexing $q(A)$ with the semantic role i, and adding $\langle i, q(A)\rangle$ to the store S. A lexical verb of arity n is assumed to denote the pair $\langle \emptyset, R\rangle$, where R is an n-ary polarity relation.

Definition 9.27 (**truth**). Let v be a lexical verb of arity n and $d_1, \ldots d_n$ be expressions denoting some w-quantifiers $[\![d_1]\!], \ldots, [\![d_n]\!]$, respectively, and let π be a permutation of $\{1, \ldots, n\}$. Then:

$$d_{\pi(1)} \bullet_{\pi(1)} (d_{\pi(2)} \bullet_{\pi(2)} (\ldots (d_{\pi(n)} \bullet_{\pi(n)} v))) \text{ is true iff}$$

$$\left\langle [\![d_1]\!], \ldots, [\![d_n]\!], + \right\rangle \in \textbf{c-exp}_{\pi(1)}([\![d_{\pi(1)}]\!], \ldots (\textbf{c-exp}_{\pi(n)}([\![d_{\pi(n)}]\!], P)))$$

The different readings of multiply quantified sentences can be computed by different orderings of i-expansion (i.e. different permutations). Consider the sentence (27) in a model in which there are two papers p_1 and p_2, two referees r_1 and r_2 and where r_1 read p_1 and r_2 read p_2, i.e., $[\![read]\!] = \{\langle r_1, p_1\rangle, \langle r_2, p_2\rangle\}$:

(27) At least one referee read every paper.

In this model, the sentence is true if the object has wide scope over the subject (for every paper p_i there is a possibly different referee r_j such that p_i was read by r_j) but false under the reading according to which the subject has wide scope over the object (since it is not the case that there is at least one referee which read every paper). First we show that the sentence is true under the object wide scope reading. Let $Q_1 = [\![\textit{at least one referee}]\!]$, $Q_2 = [\![\textit{every paper}]\!]$. We show that:

$$\langle Q_1, Q_2, + \rangle \in \textbf{s-exp}_2(Q_2, \textbf{s-exp}_1(Q_1, [\![read]\!]^*))$$

Since $\langle r_1, p_1, + \rangle \in [\![read]\!]^*$ and $[\langle r_1, p_1, + \rangle]_1^{[\![read]\!]^*} = \{r_1\}$ is a witness set of Q_1, it follows that $\langle Q_1, p_1, + \rangle \in \textbf{s-exp}_1(Q_1, [\![read]\!]^*)$. Similarly, since $\langle r_2, p_2, + \rangle \in [\![read]\!]^*$ and the set of 1-fillers $[\langle r_2, p_2, + \rangle]_1^{[\![read]\!]^*} = \{r_2\}$ is a witness set of Q_1 it follows that $\langle Q_1, p_2, + \rangle \in \textbf{s-exp}_1(Q_1, [\![read]\!]^*)$. Finally, because $[\langle Q_1, p_1, + \rangle]_2^{\textbf{s-exp}_1(Q_1, [\![read]\!]^*)} = \{p_1, p_2\}$ is a witness set of Q_2 it follows that $\langle Q_1, Q_2, + \rangle \in \textbf{s-exp}_2(Q_2, \textbf{s-exp}_1(Q_1, [\![read]\!]^*))$.

Next we show that the subject wide scope reading of (27) is false in our model, i.e. we show that:

$$\langle Q_1, Q_2, -\rangle \in \textbf{c-exp}_1(Q_1, \textbf{c-exp}_2(Q_2, [\![read]\!]^*))$$

Since $\langle r_1, p_1, +\rangle \in [\![read]\!]^*$ and $[\langle r_1, p_1, +\rangle]_2^{[\![read]\!]^*} = \{p_1\}$ is not a witness set of Q_2, it follows by the second clause of 2-expansion that $\langle r_1, Q_2, -\rangle \in \textbf{c-exp}_2(Q_2, [\![read]\!]^*)$. Similarly, since $\langle r_2, p_2, +\rangle \in [\![read]\!]^*$ and $[\langle r_2, p_2, +\rangle]_2^{[\![read]\!]^*} = \{p_2\}$ is not a witness set of Q_2, it again follows by application of clause 2 of the 2-expansion that $\langle r_2, Q_2, -\rangle \in \textbf{c-exp}_2(Q_2, [\![read]\!]^*)$. Next, note that

$$[\langle r_1, Q_2, -\rangle]_1^{\textbf{c-exp}_2(Q_2, [\![read]\!]^*)} = \{r_1, r_2\} = [\![referee]\!]$$

and that $\emptyset$ is not a witness set of Q_1. Therefore, by clause four of 1-expansion, we have that $\langle Q_1, Q_2, -\rangle \in \textbf{c-exp}_1(Q_1, \textbf{c-exp}_2(Q_2, [\![read]\!]^*))$.
We are now ready to show that, given our definition of the expansion operation above, the empty-set property of w-quantifiers is crucial for the quantificational complexity of quantified statements. Proposition 9.28 states that quantified statements with non-empty-set quantifiers allow for simpler verification algorithms than quantified statements with empty-set quantifiers.

Proposition 9.28. Let $A_1, \ldots, A_n \subseteq M$ and let $d_1, \ldots, d_n$ denote the non-empty-set w-quantifiers $q_1(A_1), \ldots, q_n(A_n)$, respectively. Let v denote the polarity relation P^* of some predicate $P \subseteq E^n$ and π be some permutation of $\{1, \ldots, n\}$. Then, the formula $d_{\pi(1)} \bullet_{\pi(1)} (d_{\pi(2)} \bullet_{\pi(2)}, \ldots, (d_{\pi(n)} \bullet_{\pi(n)} v) \ldots)$ is true iff:

$$\langle q_1(A_1), \ldots, q_n(A_n), +\rangle$$

$$\in \textbf{s-exp}_{\pi(1)}\left(q_{\pi(1)}\left(A_{\pi(1)}\right), \ldots, \textbf{s-exp}_{\pi(n)}\left(q_{\pi(n)}\left(A_{\pi(n)}\right), P^*\right) \ldots\right)$$

Proof. We limit ourselves here to sketching the proof idea for sentences with two quantifiers and linear scope. The following list illustrates the four possible ways in which the positive triple $\langle q_1(A_1), q_2(A_2), +\rangle$ could in principle have been added to P^* when expanded by $\textbf{c-exp}_1(q_1(A_1), \textbf{c-exp}_2(q_2(A_2), P^*))$.

1. $\langle a, b, +\rangle \xrightarrow{1} \langle a, q_2(A_2), +\rangle \xrightarrow{1} \langle q_1(A_1), q_2(A_2), +\rangle$
2. $\langle a, b, +\rangle \xrightarrow{2} \langle a, q_2(A_2), -\rangle \xrightarrow{3} \langle q_1(A_1), q_2(A_2), +\rangle$
3. $\langle a, b, -\rangle \xrightarrow{3} \langle a, q_2(A_2), +\rangle \xrightarrow{1} \langle q_1(A_1), q_2(A_2), +\rangle$
4. $\langle a, b, -\rangle \xrightarrow{4} \langle a, q_2(A_2), -\rangle \xrightarrow{3} \langle q_1(A_1), q_2(A_2), +\rangle$

Note that the last three sequences require the antecedent of the third expansion clause to be true. Since the antecedent of the third expansion clause requires the empty set to be a witness set of the respective w-quantifier, it follows that these three sequences cannot have applied, since q_1 and q_2 do not have the empty set as a witness set. So if the triple $\langle q_1(A_1), q_2(A_2), +\rangle$ can be added to $\textbf{c-exp}_1(q_1(A_1), \textbf{c-exp}_2(q_2(A_2), P^*))$, it can only be added by expanding some triple $\langle a, b, +\rangle \in P^*$ as illustrated in the first sequence. Crucially, this expansion sequence requires only the antecedent of the first expansion clause to be true. Consequently, the tuple $\langle q_1(A_1), q_2(A_2), +\rangle$ is an element of $\mathbf{c\text{-}exp}_1(q_1(A_1), \mathbf{c\text{-}exp}_2(q_2(A_2), P^*))$ iff it is also an element of $\mathbf{s\text{-}exp}_1(q_1(A_1), \mathbf{s\text{-}exp}_2(q_2(A_2), P^*))$.

It follows immediately from this proof that the first rule of the simple expansion operation suffices to evaluate multiply quantified sentences that do not contain any empty-set quantifiers. Please recall that this is exactly how truth evaluation was sketched informally in the beginning of Section 2.[20]

Importantly, **s-exp** does not involve polarity reversal, i.e. addition of positive information based on negative information, or *vice versa*. For example, from the positive information that $\langle a, b, +\rangle \in B^*$ we can infer the positive information that a stands in the B relation to some book (i.e. that $\langle a, some([\![book]\!]), +\rangle$ is in the expansion of B^* by $some([\![book]\!])$ at position 2). The verification of statements involving empty-set quantifiers may require polarity reversal. For example, given the negative information that a does not stand in the B relation to any cat (i.e. $\langle a, c, -\rangle \in B^*$ for every cat c), expansion of the second position by $no([\![cat]\!])$ adds the positive information that a stands in the B relation to no cat (i.e. $\langle a, no([\![cat]\!]), +\rangle$ is added by 2-expanding B^* by $no([\![cat]\!])$). If, as we assume, the full blown expansion **c-exp** is costlier than **s-exp**, it immediately follows that empty-set quantifiers should be more difficult to process than quantifiers which do not contain the empty set. As a first corollary, it follows that monotone decreasing quantifiers are generally more difficult to process than monotone increasing ones, because the monotone increasing quantifiers standardly considered are non-empty-set quantifiers, whereas all monotone decreasing quantifiers are empty-set quantifiers.

10 CONCLUSIONS AND OUTLOOK

There are a number of ways in which the present proposal needs to be extended in future work. First of all, in its present form the proposed model is limited to the analysis of quantifiers which are equivalent to type $\langle 1, 1\rangle$ quantifiers satisfying conservativity and extension. But what about quantifiers of different type, e.g., the type $\langle 1, 1, 1\rangle$ quantifier *More A than B are C* $(= |A \cap C| > |B \cap C|$ or the $\langle 1, 2\rangle$ quantifier *each other*? To be able to deal with these quantifiers the operation of *i*-expansion would have to be generalized accordingly. Secondly, by allowing for partial knowledge about predicate extensions – for instance, by having $\langle a, b, +\rangle \in P^*$ stand for the interpreter knowing that a stands in relation P to b, $\langle a, b, -\rangle \in P^*$ stand for the interpreter knowing that a does not stand in relation P to b, and $\langle a, b, ?\rangle \in P^*$ stand for the interpreter not knowing whether a stands in relation P to b – we open up a new way of accounting for superlative quantifiers like *at least n* and *at most n*, since they are typically used if the speaker does not know the precise extension of a predicate. This extension would be especially important to fully account

20 The reader may wonder why the definition of the simple expansion operation contains two additional clauses. The complete set of rules becomes only important for the evaluation of multiply quantified sentences that contain both, empty-set as well as non-empty-set quantifiers, by recursively applying a sequence of simple and complex expansion operations; cf. the sequences on the right-hand side of the table in (10). In order to keep things simple, we decided to only include the more general version of the simple expansion operation. For sentences that contain only non-empty-set quantifiers the additional two rules are, however, not necessary.

for the observed processing differences between superlative and Aristotelian quantifiers in Exp. 3. Also, as already pointed out in Section 2, the present account is restricted to scope readings of multiply quantified sentences, but has nothing to say about other readings of multiply quantified sentences such as collective readings. We have hinted at the 'reductionist' solution in Peters & Westerståhl (2006, p. 351) and Szymanik (2010) that has been proposed for cumulation. However, according to our intuitions this solution does not provide us with a fully satisfactory solutions for collective interpretations as we find them with truly collective predicates as illustrated in (28). Instead, we would rather prefer an extension of the proposed account to quantification over plural objects.

(28) Tausend Ritter umzingelten zwei Festungen.
Thousand knights encircled two castles.

Finally, and perhaps most importantly our account does not take into consideration the internal makeup of quantifier expressions. Some quantificational determiners are focus sensitive operators or involve comparative or superlative morphology – points that have featured prominently in the semantic literature on quantificational determiners. Obviously, the proposed theory has to be extended to be able to account for the insights of decompositional analyses of quantificational determiners in the spirit of Krifka (1999), Hackl (2000), Geurts & Nouwen (2007) and Hackl (2009). And last but not least, at the end of the day a cognitively realistic quantification theory should of course be spelled out in a procedural way.

Acknowledgements

This research was supported by the German Science Foundation funding the projects *Composition in Context* (CiC) of the Special Priority Program *XPrag.de* (O.B.) and project B1 *Incrementality of Semantic Interpretation* of Collaborative Research Centre 833 (O.B. and F.S.) as well as project A8 *Constructions: Bridging the gap between syntax and semantics* of Collaborative Research Centre 673 (U.K.). The authors are especially grateful to three anonymous reviewers and to the editor, Rick Nouwen, for their valuable comments on earlier versions of this paper. The usual disclaimer applies.

References

Anderson, Cathrine (2004), *The Structure and Real Time Comprehension of Quantifier Scope Ambiguity* . Ph.D. thesis, Northwestern University, Evanston, Illinois.

Augurzky, Petra, Bott, Oliver, Sternefeld, Wolfgang, & Rolf Ulrich (**2016**), 'Are all triangles blue? – ERP evidence for the incremental processing of german quantifier restriction'. *Language and Cognition* **9**: 603–36.

Barwise, John & Robin Cooper (1981), 'Generalized quantifiers and natural language'. *Linguistics and Philosophy* **4**: 159–219.

Beghelli, Filippo, Ben-Shalom, Dorit, & Anna Szabolcsi (1997), 'Variation, distributivity, and the illusion of branching'. In A. Szabolcsi (ed.), *Ways of Scope Taking*. Kluwer. Dordrecht. 29–69.

van Benthem, Johann (1986), *Essays in Logical Semantics. Reidel*. Dordrecht.

Bott, Oliver, Augurzky, Petra, Sternefeld, Wolfgang, & Rolf Ulrich (2017), 'Incremental generation of answers during the comprehension of questions with quantifiers'. *Cognition*, **166**: 328–43.

Bott, Oliver, Klein, Udo, & Fabian Schlotterbeck (2013), 'Witness sets, polarity reversal and the processing of quantified sentences'. In M. Aloni, M. Franke, and F. Roelofsen (eds.),*Proceedings of the 19th Amsterdam Colloquium*. University of Amsterdam, Amsterdam.

Bott, Lewis & Ira Noveck (2004), 'Some utterances are underinformative: the onset and time course of scalar inferences'. *Journal of Memory and Language* **51**: 437–57.

Bott, Oliver & Fabian Schlotterbeck (2012). 'Incremental truth value judgments'. In B. Stolterfoht & S. Featherston (eds.), *Empirical Approaches to Linguistic Theory*. Mouton de Gruyter, Berlin. 3–28.

Bott, Oliver & Fabian Schlotterbeck (2015), 'The processing domain of scope interaction'. *Journal of Semantics* **32**: 39–92.

Bott, Oliver & Wolfgang Sternefeld (2017), 'An event semantics with continuations for incremental interpretation'. *Journal of Semantics* **34**: 201–36.

Büring, Daniel (2008), 'The least at least can do'. In C. B. Chang and H. J. Haynie (eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*. Cascadilla Proceedings, Somerville. 114–120.

Caplan, David & Gloria S. Waters (1999), 'Verbal working memory and sentence comprehension'. *Behavioral and Brain Sciences* **22**: 77–94.

Church, Alonzo (1936), 'An unsolvable problem of elementary number theory'. *American Journal of Mathematics* **58**: 345–63.

Clark, Herbert & William Chase (1972), 'On the process of comparing sentences against pictures'. *Cognitive Psychology* **3**: 472–517.

Clark, Herbert H. (1973), 'The language-as-fixed-fallacy: a critique of language statistics in psychological research'. *Journal of Verbal Learning and Verbal Behaviour* **12**: 335–59.

Coppock, Elizabeth & Thomas Brochhagen (2013), 'Raising and resolving issues with scalar modifiers'. *Semantics & Pragmatics* **6**: 1–57.

Cummins, Chris & Napoleon Katsos (2010), 'Comparative and superlative quantifiers: pragmatic effects of comparison type'. *Journal of Semantics* **27**: 271–305.

Dehaene, Stanislas (1997). *The Number Sense: How the Mind Creates Mathematics*. Oxford University Press, Oxford.

Dehaene, Stanislav & Jean-Pierre Changeux (1993), 'Development of elementary numerical abilities – a neuronal model'. *Journal of Cognitive Neuroscience* **5**: 390–407.

Dehaene, Stanilav. & Andreas Nieder (2009), 'Representation of number in the brain'. *Annual Review of Neuroscience* **32**: 185–208.

Deschamps, Isabelle, Agmon, Galit, Loewenstein, Yonatan & Yosef Grodzinsky (2015), 'The processing of polar quantifiers, and numerosity perception'. *Cognition* **143**: 115–28.

Feigenson, Lisa, Dehaene, Stanislas, & Elizabeth Spelke (2004), 'Core systems of number'. *TRENDS in Cognitive Sciences* **8**: 307–14.

Filik, Ruth, Leuthold, Hartmut, Moxey, Linda M., & Anthony J. Sanford (2011), 'Anaphoric reference to quantified antecedents: an event-related potential study'. *Neuropsychologia* **49**: 3786–94.

von Fintel, Kai (1993) , 'Exceptive constructions'. *Natural Language Semantics* **1**: 123–48.

von Fintel, Kai & Lisa Matthewson (2008), 'Universals in Semantics'. *The Linguistic Review* **25**: 139–201.

Frazier, Lyn, Clifton, Chuck, Rayner, Keith, Deevy, P., Koh, S., & Markus Bader (2005), 'Interface problems: structural constraints on interpretation?' *Journal of Psycholinguistic Research* **34**: 201–31.

Freunberger, Dominik & Mante S. Nieuwland (2016), 'Incremental comprehension of spoken quantifier sentences: evidence from brain potentials'. *Brain Research* **1646**: 475–81.

Frey, Werner (1993), *Syntaktische Bedingungen für die Semantische Repräsentation*. Über Bindung, Implizite Argumente und Skopus. Vol. XXXV of Studia Grammatica. Akademie Verlag, Berlin.

García-Álvarez, Iván (2008), *Generality and Exception: A Study in the Semantics of Exceptives. Ph.D. thesis, Stanford University*, Stanford, CA.

Geurts, Bart (2010), *Quantity Implicatures*. Cambridge University Press, Cambridge, MA.

Geurts, Bart, Katsos, Napoleon, Moons, Jonas & Leo Noordman (2010), 'Scalar quantifiers: logic, acquisition, and processing'. *Language and Cognitive Processes* **25**: 130–48.

Geurts, Bart & Rick Nouwen (2007), ' "At least" et al.: the semantics of scalar modifiers'. *Language* **83**: 533–59.

Geurts, Bart & Rob van der Sandt (1999), 'Domain restriction'. In P. Bosch and R. van der Sandt (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge University Press, Cambridge. 268–92.

Geurts, Bart & Frans van der Slik (2005), 'Monotonicity and processing load'. *Journal of Semantics* **22**: 97–117.

Gibbs, Raymond W. Jr. (2017), 'Experimental pragmatics'. In Y. Huang (ed.), *Oxford Handbook of Pragmatics*. Oxford University Press, Oxford. 310–25.

Grice, H. Paul (1991), *Studies in the Way of Words*. Harvard University Press, Cambridge, MA.

Grodner, Daniel J., Klein, Natalie M., Carbary, Cathleen M., & Michael K. Tanenhaus (2013), '"some," and possibly all, scalar inferences are not delayed: evidence for immediate pragmatic enrichment'. *Cognition* **116**: 42–55.

Haberlandt, Karl (1994), 'Methods in reading research'. In M. A. Gernsbacher (ed.) *Handbook of Psycholinguistics (1st ed.)* . Academic Press, San Diego, CA. 1–31.

Hackl, Martin (2000), *Comparative Quantifiers* . Ph.D. thesis, MIT, Massachussetts.

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

Haskell, Todd R., Thornton, Robert, & Maryellen C. MacDonald (2010), 'Experience and grammatical agreement: statistical learning shapes number agreement production'. *Cognition* **114**: 151–64.

Heim, Irene & Angelika Kratzer (1998), *Semantics in Generative Grammar*. Wiley-Blackwell, Cambridge, MA.

Hendriks, H. (1993), *Studied Flexibility. Categories and Types in Syntax and Semantics*. Ph.D. thesis, ILLC, University of Amsterdam, Amsterdam.

Hoeksema, Jack (1990), 'Exploring exception phrases'. In M. Stokhof and L. Torenvliet (eds.), *Proceedings of the 7th Amsterdam Colloquium*. ITLI, Amsterdam. 165–90.

Hopcroft, John E. & Jeffrey D. Ullman (1979). *Introduction to Automata Theory, Languages, and Computation*. Reading: Addison Wesley. UK.

Horn, Laurence (1968), *On the Semantic Properties of Logical Operators in English*. Ph.D. thesis, UCLA. Los Angeles.

Huang, Yi Ting & Jesse Snedeker (2009), 'Online interpretation of scalar quantifiers: insight into the semantics-pragmatics interface'. *Cognitive Psychology* **58**: 376–415.

Huang, Y. T., Spelke, Elizabeth & Jesse Snedeker (2013), 'What exactly do numbers mean?' *Language Learning and Development* **9**: 105–29.

Hunter, Tim & Jeffrey Lidz (2013), 'Conservativity and learnability of determiners'. *Journal of Semantics* **30**: 315–34.

Hunter, Tim, Lidz, Jeffrey, Odic, Darko, & Alexis Wellwood (2016), 'On how verification tasks are related to verification procedures: a reply to Kotek et al.' *Natural Language Semantics* **25**: 91–107.

Johnson-Laird, Peter N. (1983), *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Harvard University Press, Cambridge, MA.

Johnson-Laird, Peter N. & Bruno G. Bara (1984), 'Syllogistic inferences'. *Cognition* **16**: 1–61.

Just, Marcel Adam & Patricia Ann Carpenter (1971), 'Comprehension of negation with quantification'. *Journal of Verbal Learning and Verbal Behaviour* **10**: 244–53.

Kaan, Edith, Dallas, Andrea C., & Christopher M. Barkley (2007), 'Processing bare quantifiers in discourse'. *Brain Research* **1146**: 199–209.

Keenan, Edward L. & Jonathan Stavi (1986), 'A semantic characterization of natural language determiners'. *Linguistics and Philosophy* **9**: 253–326.

Knoeferle, Pia, Crocker, Matthew W., Scheepers, Christoph, & Martin J. Pickering (2005), 'The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events'. *Cognition* **95**: 95–127.

Koster-Moeller, Jorie, Varvoutis, Jason & Martin Hackl (2008), 'Verification procedures for modified numeral quantifiers'. In N. Abner and J. Bishop (eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics*. Cascadilla Proceedings, Somerville. 310–17.

Kotek, Hadas, Sudo, Yasutada, & Martin Hackl (2015), 'Experimental investigations of ambiguity: the case of *most*'. *Natural Language Semantics* **23**: 119–56.

Kounios, John & Phillip J. Holcomb (1992), 'Structure and process in semantic memory: evidence from event-related potentials and reaction times'. *Journal of Experimental Psychology: General* **121**: 459–79.

Krifka, Manfred (1999), 'At least some determiners aren't determiners'. In K. Turner (ed.), *The Semantics/Pragmatics Interface from Different Points of View*. Elsevier, Oxford. 257–91.

Kutas, Marta & Kara D. Federmeier (2011), 'Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP)'. *Annual Review of Psychology* **62**: 621–47.

Levinson, Stephen C. (2000), *Presumptive Meaning*. MIT Press, Cambridge, MA.

Lidz, Jeffrey, Pietroski, Paul, Halberda, Justin, & Tim Hunter (2011), 'Interface transparency and the psychosemantics of *most*'. *Natural Language Semantics* **19**: 227–56.

Lloyd, John Wylie (1987), *Foundations of Logic Programming*. Springer, Berlin.

Marr, David (1982), *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Co., Inc., New York, NY.

Marslen-Wilson, W. (1975) , 'Sentence perception as an interactive parallel process'. *Science* **189**: 226–28.

Marty, Paul, Chemla, Emmanuel, & Benjamin Spector (2013), 'Interpreting numerals and scalar items under memory load'. *Lingua* **133**: 152–63.

McCloskey, Michael & Kurt Bigler (1980), 'Focused memory search in fact retrieval'. *Memory & Cognition* **8**: 253–64.

McMillan, Corey, Clark, Robin, Moore, Peachie, Devita, Christian, & Murray Grossman (2005), 'Neural basis for generalized quantifier comprehension'. *Neuropsychologia* **43**: 1729–37.

McMillan, Corey T., Clark, Robin, Moore, Peachie, & Murray Grossman (2006), 'Quantifier comprehension in corticobasal degeneration'. *Brain and Cognition*, **62**: 250–60.

Moltmann, Friederike (1995) , 'Exception sentences and polyadic quantification'. *Linguistics & Philosophy* **18**: 223–80.

Moschovakis, Yannis (1994), 'Sense and denotation as algorithm and value.' In J. Oikkonen and J. Väänänen (eds.), *Lecture Notes in Logic*, vol. 2. Association for Symbolic Logic, Ltd., Natick. 210–49.

Mostowski, Marcin (1998) , 'Computational semantics for monadic quantifiers'. *Journal of Applied Non-Classical Logics* **8**: 107–21.

Moxey, Linda M. (2006), 'Effects of what is expected on the focussing properties of quantifiers: a test of the presupposition-denial account'. *Journal of Memory and Language* **55**: 422–39.

Moxey, Linda M. & Anthony J. Sanford (1987), 'Quantifiers and focus'. *Journal of Semantics* **5**: 189–206.

Moxey, Linda M. & Anthony J. Sanford (1993), *Communicating Quantities: A Psychological Perspective. Lawrence Erlbaum Associates*, Hove, UK.

Moxey, Linda M., Sanford, Anthony J., & E. Dawydiak (2001), 'Denials as controllers of negative quantifier focus'. *Journal of Memory and Language* **44**: 427–42.

Moyer, Robert S. & Thomas K. Landauer (1967), 'Time required for judgements of numerical inequality'. *Nature* **215**: 1519–20.

Nieder, Andreas (2016a) , 'Representing something out of nothing: the dawning of zero'. *Trends in Cognitive Sciences* **20**: 830–42.

Nieder, Andreas (2016b), 'The neural code for number'. *Nature Reviews Neuroscience*, **17**: 366–82.

Nieder, Andreas, Freedman, David J., & Earl K. Miller (2002), 'Representation of the quantity of visual items in the primate prefrontal cortex.' *Science* **297**: 1708–11.

Nieuwland, Mante S. (2016), 'Quantification, prediction, and the online impact of sentence truth-value: evidence from event-related potentials'. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **42**: 316–34.

Nouwen, Rick (2003), 'Complement anaphora and interpretation'. *Journal of Semantics* **20**: 73–113.

Nouwen, Rick (2010), 'Two kinds of modified numerals'. *Semantics & Pragmatics* **3**: 1–41.

Olm, Christopher A., McMillan, Corey T., Spotorno, Nicola, & Murray Grossman (2014), 'The relative contrbutions of frontal and parietal cortex for generalized quantifier comprehension.' *Frontiers in Human Neuroscience* **8**: 1–8.

Paterson, Kevin B., Filik, Ruth, & Linda M. Moxey (2009), 'Quantifiers and discourse processing'. *Language and Linguistics Compass* **3**: 1390–402.

Paterson, Kevin B., Sanford, Anthony J., Moxey, Linda M., & E. Dawydiak (1998), 'Quantifier polarity and referential focus during reading'. *Journal of Memory and Language* **39**: 290–306.

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

Piantadosi, Steven, Goodman, Noah, Ellis, Benjamin, & Joshua Tenenbaum (2008), 'A Bayesian model of the acquisition of compositional semantics'. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. 1620–5. Cognitive Science Society, Washington, DC.

Pietroski, Paul, Lidz, Jeffrey, Hunter, Tim, & Justin Halberda (2009), 'The meaning of 'most': semantics, numerosity, and psychology'. *Mind and Language* **24**: 554–85.

Radó, Janina & Oliver Bott (2012), 'Underspecified representations of scope ambiguity?' In M. Aloni, V. Kimmelman, F. Roelofsen, G. Sassoon, K. Schulz, and M. Westera (eds.), *Logic, Language and Meaning. Volume 7218 of Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg. 180–9.

Ramirez-Cardenas, Araceli, Moskaleva, Maria, & Andreas Nieder (2016), 'Neuronal representation of numerosity zero in the primate parieto-frontal number network'. *Current Biology* **26**: 1285–94.

Reder, Lynne M. (1982), 'Strategy selection in question answering'. *Cognitive Psychology* **19**: 90–138.

Reyle, Uwe (1993) , 'Dealing with ambiguities by underspecification: construction, representation and deduction'. *Journal of Semantics* **10**: 123–79.

Robaldo, Livio, Szymanik, Jakub, & Ben Meijering (2014), 'On the identification of quantifiers' witness sets: a study of multi-quantifier sentences'. *Journal of Logic, Language and Information* **23**: 53–81.

Rooth, Mats (1985), *Association with Focus* . Ph.D. thesis, University of Massachusetts, Amherst.

Rooth, Mats (1992) , 'A theory of focus interpretation'. *Natural Language Semantics* **1**: 75–116.

Sanford, Anthony J., Dawydiak, E., & Linda M. Moxey (2007), 'A unified account of quantifier perspective effects in discourse'. *Discourse Processes* **44**: 1–32.

Sanford, Anthony J., Moxey, Linda M. & Kevin Paterson (1994), 'Psychological studies of quantifiers'. *Journal of Semantics* **10**: 153–70.

Sanford, Anthony J., Moxey, Linda M. & K.B. Paterson (1996), 'Attentional focussing with quantifiers in production and comprehension.' *Memory & Cognition* **24**: 144–55.

Sanford, Anthony J. & Patrick Sturt (2002), 'Depth of processing in language comprehension: not noticing the evidence'. *Trends in Cognitive Sciences* **6**: 382–6.

Sanford, Anthony J., Williams, C., & N. Fay (2001), 'When being included is being excluded: a note on complement set focus and the inclusion relation'. *Memory & Cognition* **29**: 1096–101.

Schlotterbeck, Fabian (2017), *From Truth Conditions to Processes: How to Model the Processing Difficulty of Quantified Sentences Based on Semantic Theory. Ph.D. thesis*, Eberhard Karls Universität, Tübingen.

Schwarz, Bernhard (2013), ''at least' and quantity implicature: choices and consequences'. In M. Aloni, M. Franke, and F. Roelofsen (eds.), *Proceedings of the 19th Amsterdam Colloquium*. University of Amsterdam, Amsterdam.

Schwarz, Bernhard (2016), 'At least and ignorance: a reply to Coppock & Brochhagen 2013'. *Semantics & Pragmatics* **9**: 1–17.

Singer, Murray (1984) , 'Towards a model of question answering: yes-no questions'. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **10**: 285–97.

Spychalska, Maria (2018) , 'At least not false, at most possible: between truth and assertability of superlative quantifiers'. *Synthese* **195**: 571–602.

Steinert-Threlkeld, Shane (2016) , 'Some properties of iterated languages'. *Journal of Logic, Language and Information* **25**: 191–213.

Steinert-Threlkeld, Shane & Thoams F. Icard, III (2013), 'Iterating semantic automata.' *Linguistics and Philosophy* **36**: 151–73.

Sternefeld, Wolfgang (2006), *Syntax, vol. 1*. Stauffenberg, Tübingen.

Szabolcsi, Anna (2010), *Quantification*. Cambridge University Press, Cambridge, MA.

Szymanik, Jakub (2007), 'A comment on a neuroimaging study of natural language quantifier comprehension'. *Neuropsychologia* **45**: 2158–60.

Szymanik, Jakub (2009). *Quantifiers in Time and Space - Computational Complexity of Generalized Quantifiers in Natural Language. ILLC Dissertation Series. Institute for Logic, Language and Computation*, Amsterdam.

Szymanik, Jakub (2010) , 'Computational complexity of polyadic lifts of generalized quantifiers in natural language'. *Linguistics and Philosophy* **33**: 215–50.

Szymanik, Jakub (2016), *Quantifiers and Cognition. Logical and Computational Perspectives*. Springer, Cham.

Szymanik, Jakub & Marcin Zajenkowski (2010), 'Comprehension of simple quantifiers: Empirical evaluation of a computational model'. *Cognitive Science* **34**: 521–32.

Szymanik, Jakub & Marcin Zajenkowski (2011), 'Contribution of working memory in parity and proportional judgments'. *Belgian Journal of Linguistics* **25**: 176–94.

Szymanik, Jakub & Marcin Zajenkowski (2013), 'Monotonicity has only a relative effect on the complexity of quantifier verification'. In M. Aloni, M. Franke, and F. Roelofsen (eds.), *Proceedings of the 19th Amsterdam Colloquium*. 219–25. University of Amsterdam, Amsterdam.

Tomaszewicz, Barbara M. (2011), 'Verification strategies for two majority quantifiers in Polish'. In I. Reich, E. Horch, and D. Pauly, D. (eds.), *Proceedings of Sinn und Bedeutung 15*. Saarland University Press. Saarbrücken. 114–20.

Troiani, Vanessa, Peelle, Jonathan E., Clark, Robin, & Murray Grossman (2009), 'Is it logical to count on quantifiers? Dissociable neural networks underlying numerical and logical quantifiers'. *Neuropsychologia* **47**: 104–11.

Tunstall, Susanne Lynn (1998), *The Interpretation of Quantifiers: Semantics and Processing* . Ph.D. thesis, University of Massachusetts, Amherst.

Urbach, Thomas P., DeLong, Katherine A., & Marta Kutas (2015), 'Quantifiers are incrementally interpreted in in context, more than less'. *Journal of Memory and Language* **83**: 79–96.

Urbach, Thomas P. & Marta Kutas (2010), 'Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation'. *Journal of Memory and Language* **63**: 158–79.

Verguts, Tom & Wim Fias (2004), 'Representation of number in animals and humans: a neural model'. *Journal of Cognitive Neuroscience* **16**: 1493–504.

Wei, W., Chen, C., Yang, T., Zhang, H., & X. Zhou (2012), 'Dissociated neural correlates of quantity processing of quantifiers, numbers, and numerosities'. *Human Brain Mapping* **35**: 444–54.

Wiese, Heike (2003), *Numbers, Language, and the Human Mind*. Cambridge University Press, Cambridge, UK.

Wijnen, Frank & Edith Kaan (2006), 'Dynamics of semantic processing: the interpretation of bare quantifiers'. *Language and Cognitive Processes* **21**: 684–720.

Zajenkowski, Marcin, Styła, Rafał, & Jakub Szymanik (2011), 'A computational approach to quantifiers as an explanation for some language impairments in schizophrenia'. *Journal of Communication Disorders* **44**: 595–600.

Zajenkowski, Marcin & Jakub Szymanik (2013), 'MOST intelligent people are accurate and SOME fast people are intelligent'. *Intelligence* **41**: 456–66.

Zajenkowski, Marcin, Szymanik, Jakub, & Maria Garraffa (2014), 'Working memory mechanism in proportional quantifier verification'. *Journal of Psycholinguistic Research* **43**: 839–53.