THE ACQUISITION OF PLURAL EXPRESSIONS

How children learn to navigate the logical space in the realm of pluralities

By

Magdalena Roszkowski

Submitted to
Central European University
Department of Cognitive Science

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Cognitive Science

Supervisor: Ernő Téglás
Secondary Supervisor: György Gergely

Vienna, Austria

2025

Declaration of Authorship

I, the undersigned, Magdalena Roszkowski, candidate for the PhD degree in Cognitive Science, declare herewith that the present thesis is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright. I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Vienna, 30 April 2025

_____

Magdalena Roszkowski

Copyright Notice

Abstract

A longstanding and ever-fascinating puzzle is how children acquire language so rapidly and seemingly effortlessly, raising questions about the cognitive prerequisites that enable this process and the expectations that guide it. This dissertation investigates the acquisition of various constructions that involve plural expressions and the related notions of distributivity and cumulativity as well as homogeneity and (non-)maximality. It explores questions concerning children's early representational capacities, the relationship between linguistic representations and conceptual development and principles that support the mapping between language and the world. Specifically, it asks how children come to form complex semantic representations, which expectations about meaning they bring to the learning task and how the acquisition of expressions that encode plurality relates to more general abilities of reasoning about multitudes. The first part examines how children acquire abstract meaning representations in the domain of pluralities and addresses some fundamental questions regarding the acquisition of functional elements. By using a structural priming paradigm we investigate whether preschool-aged children are able to represent cumulative and distributive meanings of ambiguous plural sentences, even before they have mastered the truth-conditions of distributive universal quantifiers. The findings reveal priming effects for both cumulative and distributive interpretations in the absence of overt disambiguating elements, suggesting that children may have available certain logical representations prior to showing adult-like competence with the corresponding lexical items. The second part focuses on the relationship between language and thought and explores the possibility that the conceptual repertoire relevant for quantification is available at an early age. In particular, we investigate whether already preverbal infants are able to deploy the concept of exhaustivity, a notion that may later play a role in the acquisition of universal quantifiers. We present an eye-tracking study which involves a task that allows infants to learn a rule based on the quantificational properties of scenes featuring multiple agents. The results of

this inquiry are inconclusive, leaving the question open of whether the concept of exhaustivity is available preverbally. The third part studies how children navigate uncertainty in the application of linguistic expressions. We examine how preschoolers interpret definite singular and plural expressions in scenarios that involve non-maximal and heterogeneous referents through a truth-value judgment task. The findings show that children are receptive to both types of violations, indicating an early sensitivity to the vague nature of language and gaps in the extension of natural language expressions. Taken together the results provide evidence for an early ability to form complex semantic representations and to deal with the uncertainty accompanying linguistic expressions, while also pointing to potential differences between linguistic and non-linguistic representations.

Acknowledgements

Contents

Chapter 1

General introduction

Natural language provides a variety of means to convey information about multitudes or collections of entities. In languages such as English those means include definite plural phrases (1a), indefinite plural phrases (1b) and conjoined phrases (1c), all of which in principle can be used to describe situations that involve multiple individuals – here the situation in which the four dogs, Angel, Bella, Cooper and Daisy, are barking.[1]

(1) a. ***The dogs*** *are barking.*

b. ***(Four) dogs*** *are barking.*

c. ***Angel, Bella, Cooper and Daisy*** *are barking.*

Children have been shown to acquire the meaning of plural expressions, i.e. to distinguish between expressions that denote single entities and those that encode plurality, early in development, with some research suggesting that the singular/plural distinction is acquired at 24 months of age (e.g. Kouider et al., 2006; Barner et al., 2007; Wood et al., 2009; Clark and Nikitina, 2009; Davies et al., 2017 on comprehension in English-speaking children, and e.g. Brown, 1973; Park, 1978; Köpcke, 1998; Szagun, 2001; Clark and Nikitina, 2009; Haslinger and Schmitt, 2017 on spontaneous and elicited production in English- and German-speaking

---
[1]There are other ways to encode such information within and across languages and not all languages exhibit the range of means which is available in English (see e.g. Corbett, 2000; Harbour, 2014; Nouwen, 2016). Additionally, there is ongoing debate in the literature regarding the exact denotation of plurals (cf. Sauerland et al., 2005). As these topics fall outside the scope of the current work, they are not directly addressed here.

children). Around their second birthday, children start to react differently when addressed with sentences that contain singular terms and sentences containing plural terms, e.g. upon request they provide a single object or multiple objects depending on the number marking in the request (Wood et al., 2009), and to use different forms when talking about multiple objects versus single objects (though not necessarily the correct forms, see e.g. Clark and Nikitina, 2009). However, mastering plural expressions in an adult-like manner requires much more. Children must not only be sensitive to this conceptual distinction and its grammatical relevance and acquire the corresponding forms in their native language(s), they must also develop an understanding of the subtle semantic and pragmatic peculiarities of different plural expressions in various linguistic contexts, which particularly involves learning to resolve structural ambiguities and to manage referential uncertainty.

One such ambiguity to be resolved occurs systematically in sentences that contain more than one plural expression like (2), which contains two numeral-modified indefinite plurals. Such sentences allow for a distributive interpretation (2a), under which the members of the subject plurality individually have the property expressed by the predicate as in SCENARIO 1, and for a non-distributive cumulative interpretation (2b), under which the properties of the individual members sum up to the property of the whole plurality as in SCENARIO 2 (cf. Scha, 1981; Link, 1983; Krifka, 1986, a.o.).[2]

(2) *Two dogs have three toys.*

    a. **Distributive interpretation:** Two dogs *each* have three toys.

    b. **Cumulative interpretation:** Two dogs, *between them*, have three toys.

---

[2]Potentially other interpretations of such sentences are also permitted, for instance a collective interpretation under which two dogs *together* have three toys (cf. Landman, 1989, a.o.). Such an interpretation is also non-distributive in the sense that it does not involve the distribution of the property expressed by the predicate to individual dogs, but rather attributes the property to the entire group of dogs. We focus on the cumulative/distributive ambiguity, but where relevant, collective and cumulative interpretations are subsumed under the term 'non-distributive' interpretations.

(3)   a.   SCENARIO 1: Angel has three toys and Bella has three toys.

       b.   SCENARIO 2: Angel has two toys and Bella has one toy.

Referential uncertainty may arise from sentences that contain definite plurals (4). Definites have been observed to give rise to so-called homogeneity effects in that they seem to require that the set they refer to is homogeneous with respect to the property expressed in the sentence (cf. Fodor, 1970; Löbner, 1987; Schwarzschild, 1994; Križ, 2015, a.o.). While the sentence in (4a) appears to be only true in a scenario in which *all* of the relevant dogs are barking (SCENARIO 3), i.e. to be interpreted universally, the sentence in (4b) appears to be only true in a scenario in which none of the dogs is barking (SCENARIO 5). Thus, (4b) apparently is not interpreted as the negation of (4a), but rather in terms of a negated existentially quantified statement requiring that it is not true that *any* of the dogs is barking. The affirmative statement and its negative counterpart seemingly do not denote complementary sets of situations – the intermediate situation in which some but not all dogs are barking (SCENARIO 4) does not seem to be included.

(4)   a.   *The dogs are barking.*

       b.   *The dogs aren't barking.*

(5)   a.   SCENARIO 3: Angel is barking, Bella is barking, Cooper is barking and Daisy is barking.

       b.   SCENARIO 4: Angel is barking and Bella is barking, Cooper is not barking and Daisy is not barking.

       c.   SCENARIO 5: Angel is not barking, Bella is not barking, Cooper is not barking and Daisy is not barking.

At the same time, definite plurals have been argued to allow for exceptions under certain circumstances, i.e. to allow for non-maximal construals (e.g. Brisson, 1998; Lasersohn, 1999; Malamud, 2012; Križ, 2015). Consider again the example in (4a), this time uttered in the context below. It seems that given this context, (4a) may be felicitous in the non-maximal SCENARIO 6, where only three out of the four dogs are barking.

3

(6) a. CONTEXT: John goes to bed and leaves his windows open. Early in the morning he angrily shouts:

    b. *I can't sleep anymore, the dogs are barking!*

    c. SCENARIO 6: Angel is barking, Bella is barking, Cooper is barking, Daisy is still sleeping and not barking.

In the following we examine the acquisition of pluralities in the individual domain, which intuitively correspond to groups or collections of individuals, along with the related notions of cumulativity and distributivity on the one hand, and (non-)maximality and homogeneity on the other hand, thereby attempting to gain new insights into the general processes that guide the acquisition of abstract concepts and the cognitive principles mediating the relationship between language and the world.

Chapter 2 is concerned with theories of language acquisition and the question of how children build up their functional vocabulary. When acquiring the meanings of logical expressions such as *some*, *all*, *and*, and *or*, young language learners face a particularly challenging task since the acquisition of such abstract meanings is specifically constrained by the poverty of the stimulus (cf. Goodman, 1955). While some theories assume that children come to this task equipped only with a small and restricted set of conceptual primitives before constructing full-fledged adult-like logical representations when learning the corresponding linguistic expressions (e.g. Carey, 2009, 2011; Xu, 2019; Leahy and Carey, 2020), an alternative view holds that already infants possess a rich mental logical vocabulary onto which newly learned expressions are eventually mapped during language acquisition (e.g. Fodor, 1975; Macnamara, 1986; also Crain, 2012). We take the cumulative/distributive ambiguity of plural sentences as a test case and ask whether children learn to distinguish these interpretations only when they have acquired overt markers with corresponding meanings such as *each*, or whether the distinction is mentally represented before the mastery of such expression. The aim of our investigation is to find out whether the acquisition of linguistic expressions is a pre-condition for the development of abstract representations in this domain or whether there is evidence that existing knowledge structures are exploited in the course of acquisition.

4

In chapter 3, we take a further step in this direction by exploring the relationship between early conceptual development and linguistic representations. Some research suggests that already preverbal infants possess quite abstract representations in that they may for instance deploy relational concepts such as SAME (e.g. Hochmann et al., 2016, 2018; Hochmann, 2022) and derive disjunctive inferences (Cesana-Arlotti et al., 2018, 2020; Ekramnia et al., 2021; Bohus et al., 2023). With the help of a rule-learning task we study whether preverbal infants have access to basic mental building blocks underlying the cumulative/distributive distinction. In particular, we test whether infants are able to learn a rule based on quantificational properties when encoding events that involve actions performed by multiple agents. As such, our study may contribute to determining whether linguistic and non-linguistic representations share core formal properties, while also shedding some light on the questions posed by competing developmental accounts.

Chapter 4 focuses on the question of how children deal with linguistic expressions that potentially give rise to uncertainty and which expectations about the mapping between language and the world are implied by their behavior. In this line of research, we examine children's interpretations of definite descriptions. We are interested in how intermediate cases for such expressions are handled and to what extent the specific properties of those cases may affect judgments. Although previous studies have shown that, in contrast to adults, children tend to be more tolerant toward such cases when interpreting plural definites and often do not exhibit adult-like sensitivity to homogeneity violations (e.g. Karmiloff-Smith, 1979; Caponigro et al., 2012; Simon-Pearson and Syrett, 2018; Moyer et al., 2019; Tieu et al., 2019), our survey offers the first systematic investigation that targets not only plural but also singular definite descriptions. Definite singulars have likewise been argued to give rise to homogeneity effects (cf. Löbner, 2000; Križ, 2015, 2016; Paillé, 2022) and here we examine both affirmative and negative contexts to assess whether a general developmental pattern underlies the acquisition of homogeneous interpretations across expression types. Our goal is to learn about children's comprehension of definite descriptions and the developmental trajectory that leads them to adult-like meaning representations.

Taken together the findings of this thesis can enhance our understanding of the challenges of language learning in the domain of pluralities and illuminate fundamental aspects of human communication.

Chapter 2

The acquisition of distributive universal quantifiers

2.1 Introduction

Historically, quantification has always been a central topic in the study of the mind and there has been considerable interest in the acquisition of quantifying expressions such as *all*, *each*, *every* and *some* at least since Inhelder and Piaget's work on class inclusion (1958; 1964), continuing with research on children's interpretation of quantified sentences by e.g. Roeper and de Villiers (1993) and Philip (1995), and Brooks and Brain's investigation of semantic distinctions among universals (1996), with the bulk of evidence suggesting a divergence between children's and adults' interpretation of quantifying expressions. Within psychology this interest mainly stems from the idea that quantifying expressions are used to verbalize logical concepts and thus may reflect logical reasoning processes, such as deduction, and provide insights into human reasoning capacities more generally, an idea that underlies theories which assume that humans possess logically structured thoughts, either in the form of a mental logic, i.e. an inferential system based on logical connectives that correspond to natural language expressions (Braine, 1978), or a general, language-like thought device (Fodor, 1975). In linguistics, quantifiers have

7

occupied a prominent role not only because they come with complex, (nearly) universal and well-defined logical properties (cf. Barwise and Cooper, 1981; von Fintel and Matthewson, 2008) which allow for formulating exact predictions about truth-conditions and entailment patterns of sentences containing them, but also because they exhibit a particular behavior in that they give rise to ambiguities in interaction with other expressions and trigger scalar implicatures in certain environments, all of which makes them an interesting object for formal and experimental investigations and, what is more, has led to a radical change in the view of the syntax-semantics interface and the entire field (cf. Partee, 2011).

So how do children acquire the meaning of quantifying expressions and, more generally, the meanings of logical expressions? And which role does language acquisition play for the capacity to deploy the corresponding concepts in thought? It is apparent that, apart from syntactic knowledge, semantic knowledge is similarly underdetermined by the data and subject to the poverty of the stimulus, particularly for expressions whose meanings are abstract and cannot be directly observed (Goodman, 1955; Quine, 1960; Chomsky, 1965; see also Gleitman, 1990; Gleitman et al., 2005; Crain, 2012).

According to one view, most famously advanced by Fodor (1975), humans are endowed with an internal system of representations, the Language of Thought (LOT), which resembles natural language in being symbolic, combinatorial and logically structured. This system is thought to comprise logical concepts, such as quantifiers and other logical operators, and is believed to support thinking from early on, existing independently of and prior to linguistic representations. However, during language acquisition children may exploit those representational resources such that the acquisition process, which is mediated by hypothesis testing, becomes a matter of mapping between newly learned forms and pre-existing mental representations. Therefore we will refer to such theories as *Mapping Theories* and the claim that the process of acquiring a functional vocabulary involves deployment of existing mental structures as *Mapping Hypothesis*. Although assumptions about the specifics of this mapping process may differ across theories, a common view is that there is no qualitative difference between children's and adults' logical representations, and that the LOT possesses at least equal or greater expressive

power than natural language (cf. Fodor, 1975; Macnamara, 1986; see also Crain, 2012 for an alternative perspective on logical nativism).

Alternative accounts hold that, while young infants have access to sophisticated conceptual representations largely grounded in core knowledge domains, these representations are simpler in format and it is mostly through language acquisition that children develop the capacity to form general, abstract symbolic representations (cf. Carey, 2009, 2011). Thus, infants' conceptual repertoire is assumed to differ from that of adults both quantitatively and qualitatively, with language acquisition playing a crucial role in conceptual development by enriching the conceptual system and enhancing its expressive power. On this view, the acquisition of elements of the functional vocabulary involves learning specific linguistic forms and, through mechanisms such as bootstrapping, constructing corresponding abstract representations (Carey, 2009, 2011; Xu, 2019; Leahy and Carey, 2020). Within this approach it has been argued for instance that the capacity to form quantified representations and to deploy modal concepts such as possibility and necessity depend on explicit linguistic representations (cf. Carey, 2009, 2011; Leahy and Carey, 2020). We will henceforth refer to such theories as *Learning Theories* and to the assumption that the acquisition of functional vocabulary involves conceptual construction as the *Learning Hypothesis*.

Children begin producing universal quantifiers such as *every* as early as age 2 (Rasin and Aravind, 2021), however, research suggests that certain aspects of quantifying expressions remain challenging beyond the preschool years. Starting with the observation that preschoolers struggle with the correct interpretation of sentences containing a universal by Inhelder and Piaget (1964), who attributed children's errors to the absence of mental logical operations corresponding to quantified expressions during early developmental stages, later studies found that a persistent error in children's interpretation of transitive universally quantified sentences is that they seem to interpret those exhaustively on both the subject and the object (a phenomenon known as *quantifier spreading* or *exhaustive pairing*,Takahashi, 1991; Roeper and de Villiers, 1993; Philip, 1995; Drozd, 1996 a.o.).[1] In particular, children have been observed to give non-adult

---

[1]The error has also been documented to occur with intransitive sentences (e.g. children give negative answers to the question *Is every cat waving?* in scenarios where three cats are waving and and extra animal is not waving),

responses when asked questions such as (7), in that they agree when presented with depictions of scenarios that involve an exhaustive pairing of the elements denoted by the subject and the object noun phrases such as (8a), but tend to disagree in scenarios that involve an extra object such as (8b).

(7) *Is every girl riding a bike?*

(8) a. EXHAUSTIVE PAIRING: There are three girls and three bikes. Each girl is riding a different bike.

  b. NON-EXHAUSTIVE PAIRING: There are three girls and three bikes. Each girl is riding a different bike. Additionally, there is fourth bike that nobody is riding.

While many researchers have explained this error in terms of a non-adult semantic representation of the universal (e.g. Roeper and de Villiers, 1993; Philip, 1995; Drozd, 2001; Geurts, 2003), others have argued that it arises from pragmatic factors influencing children's responses in such settings (e.g. Crain et al., 1996; Aravind et al., 2017). And indeed, further studies have provided evidence in favor of the latter view, demonstrating that modifications to the experimental procedure can reduce quantifier spreading errors, e.g. changing the testing context (Crain et al., 1996), manipulating the salience of extra objects (Philip, 2011) or using more naturalistic stimuli (É. Kiss and Zétényi, 2017). Additional support for this interpretation comes from findings that even adults occasionally produce quantifier spreading errors (Brooks and Sekerina, 2006; Aravind et al., 2017).

Another commonly observed error is that children often struggle to restrict the meaning of purely distributive universals such as *each*. In adult grammar, *each* enforces a distributive meaning, whereas *all* is compatible with a collective interpretation, as evidenced by the fact that the latter can combine with collective predicates such as *gather* (9a) but the former cannot (9b) (cf. Vendler, 1962).

---

though to a lesser extent (Philip, 1995).

(9)   a.   *All boys gathered in the hall.*

       b.   * *Each boy gathered in the hall.*

Thus, the potentially ambiguous sentence in (10b) containing *each* is judged true in the distributive scenario in (11b) where each men is building his own boat, but false in the collective scenario in (11a) where the men are building a boat together. By contrast, the sentence containing the universal *all* (10a) is compatible with both distributive and collective interpretations, making it acceptable in either scenario.

(10)   a.   *All of the men are building a boat.*

       b.   *Each man is building a boat.*

(11)   There are three men.

       a.   COLLECTIVE SCENARIO: The three men are building one boat together.

       b.   DISTRIBUTIVE SCENARIO: The three men are each building a different boat.

Several studies suggest that children are not yet fully aware of the semantic distinction between *all* and *each* and do not associate a distributive meaning with the latter (e.g. Brooks and Braine, 1996; Brooks et al., 2001; Pagliarini et al., 2012; Syrett and Musolino, 2013), raising questions about the underlying source of this developmental gap.

2.2   Ambiguous plural sentences

Interestingly, even sentences that do not contain overt distributive universals like *each* may require access to a distributive quantifier meaning, in particular, ambiguous sentences that involve multiple plural expressions and allow for more than one interpretation like the sentence in (2), repeated in (12). This sentence may receive a distributive interpretation (12a), under which two dogs each have three toys (there are six toys in total), and it allows for a cumulative interpretation (12b), where two dogs have three toys between them (there are three toys in total).

(12)   *Two dogs have three toys.*

11

a. **Distributive interpretation:** Two dogs *each* have three toys.

b. **Cumulative interpretation:** Two dogs, *between them*, have three toys.

That these are truly two distinct interpretations is shown by the fact that by making the *exact* interpretation of the object numeral explicit as in (13), the sentence on its cumulative interpretation is true in a cumulative scenario but false in the distributive scenario, whereas on its distributive interpretation it is true in the distributive but false in a cumulative scenario (cf. Figure 2.1).

(13)   *Two dogs have exactly three toys.*
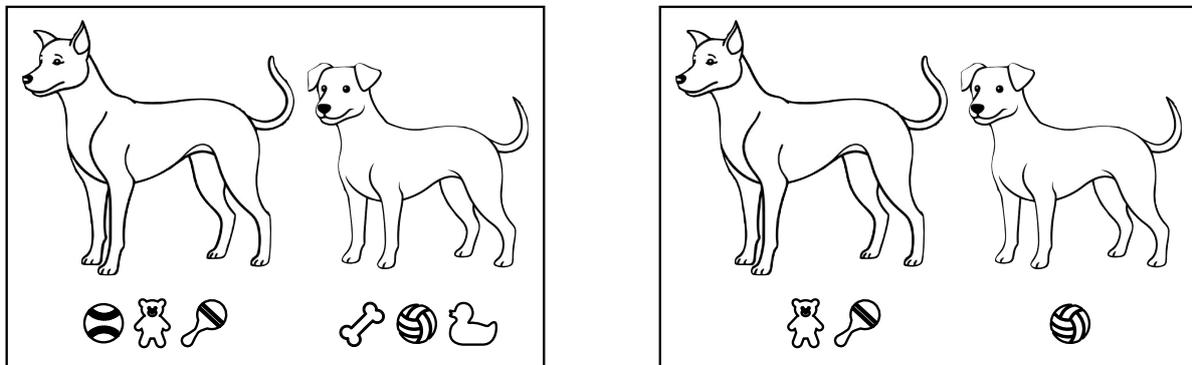
Figure 2.1: Distributive scenario in which two dogs each have three toys (left) and cumulative scenario in which two dogs between them have three toys (right)

Following Link (1987), many semantic theories assume that these different interpretations are due to structural differences at an abstract level of representation (Roberts, 1987; Schwarzschild, 1996; Lasersohn, 1998; Brisson, 2003, a.o.; see Champollion, 2020 for an overview). On Link's account, the distributive interpretation of (12) is thought to be structurally more complex and derived from the more basic, cumulative interpretation through insertion of an additional operator, the *distributivity operator* (D).[2]

(14)   a. **Cumulative structure:** [[Two dogs] [have three toys]]

b. **Distributive structure:** [[Two dogs] [**D** [have three toys]]]

The distributivity operator attaches to the predicate and functions as a universal quantifier over

---

[2]Note that some accounts assume that the distributive and the cumulative interpretation are equally complex (e.g. Sternefeld, 1998; Beck and Sauerland, 2000).

individuals, just like a silent *each*.[3] This approach is based on the idea that plural expressions such as *(the) (two) dogs* and *Angel and Bella* denote plural individuals or pluralities. Plural individuals are formed by a sum operation ($\oplus$) and stand in a part-whole relation ($\leq$) to their subparts. The part-whole relation ensures that there is a principled connection between atomic individuals, e.g. the individuals Angel and Bella, and the corresponding plural individual, e.g. the plural individual Angel$\oplus$Bella, such that if *Angel is barking* is true and *Bella is barking* is true, then it will also hold that *Angel and Bella are barking* (cf. Link, 1983). Link's approach can capture the systematic ambiguities that arise in sentences in which a plural expression combines with an ambiguous predicate as in (12):[4] The cumulative interpretation involves applying the predicate *have three toys* to the plurality consisting of two dogs, while, due to the distributivity operator which introduces a universal quantifier over atomic individuals, the distributive interpretation involves applying the predicate to each individual member making up the plurality, resulting in the truth-conditions paraphrased below.

(15)   a.   **Cumulative truth-conditions:** There is a plurality consisting of two dogs and there is a plurality consisting of three toys such that every individual dog has at least one individual toy and every individual toy is owned by at least one individual dog.

       b.   **Distributive truth-conditions:** There is a plurality consisting of two dogs such that each individual dog owns a plurality consisting of three toys.

Some experimental evidence corroborates the psychological reality of these abstract linguistic structures. For instance, studies have shown that cumulative and distributive interpretations of ambiguous plural sentences, which are similar to the example in (12), bare of any overt markers enforcing either reading, can be primed in adults (Maldonado et al., 2017; Jiang et al., 2022; also Maldonado et al., 2019).

---

[3]The systematic occurrence of plural ambiguities has been documented not only in English but also in many other languages. But while in English the distributive interpretation of sentences like (12) is available even if no overt distributive quantifier such as *each* is present, some other languages require its realization, i.e. in some languages a sentence like (12) will only receive a distributive interpretation if it contains a phonologically realized distributivity marker (see e.g. Flor et al., 2017).

[4]Some predicates are unambiguous in this respect, i.e. there are predicates that necessarily apply to atomic individuals, distributive predicates like e.g. *sleep*, and others that can only be true of plural individuals, collective predicates like e.g. *meet* (for further specifications see e.g. Landman, 1989).

Maldonado and colleagues (2017) conducted a priming study in which participants matched sentences with pictures to investigate whether cumulative and distributive interpretations of sentences containing two plural expressions such as (16) induce structural priming effects in adults, and thus whether the assumed structural differences, illustrated in (17) and (18) for (16), lead to distinct activation patterns.

(16)   *Two squares are connected to three circles.*

(17)   a.   **Cumulative structure:** [[Two squares] [are connected to [three circles]]]

      b.   **Interpretation:** Two squares, between them, are connected to three circles.

(18)   a.   **Distributive structure:** [[Two squares] [**D** [are connected to [three circles]]]]

      b.   **Interpretation:** Two squares each are connected to three circles.

The study consisted of three experiments that tested whether accessing a cumulative or a distributive interpretation of an ambiguous sentence influences the interpretation of a subsequent sentence, i.e. whether the activation of a certain representation facilitates its re-activation in a later trial.

The first experiment included three conditions, a distributive, a cumulative and a baseline condition. Across conditions, test trials consisted of two prime trials followed by one target trial. Target trials presented an ambiguous sentence containing two plural expressions (e.g. *Two circles are connected to two hearts*) alongside two pictures, one depicting a cumulative scenario (e.g. two circles of which each was connected to one heart) and another one depicting a distributive scenario (e.g. two circles of which each was connected to two hearts). Distributive and cumulative prime trials involved similar ambiguous sentences (e.g. *Two squares are connected to three circles*) which were accompanied by two pictures, one showing a scenario that renders the sentence false on any reading (e.g. two squares of which each was connected to one circle) and, depending on the condition, one that shows a distributive scenario (e.g. two squares of which each was connected to three circles) or a cumulative scenario (e.g. two squares of which one was connected to one, and the other one to two circles). In baseline prime trials, participants were presented with sentences that do not exhibit the same ambiguity (e.g. *A circle is connected*

to two squares) and two pictures of which one rendered the sentence true (e.g. one circle which was connected to two squares) and the other one false (e.g. one circle which was connected to three squares).[5] Thus, while prime trials presented one compatible and one incompatible scenario, target trials required participants to choose between two compatible scenarios.

While baseline primes were designed to enforce a neutral interpretation regarding the cumulative/distributive contrast, cumulative and distributive primes were intended to bias participants toward one interpretation. The key question was whether exposure to a particular reading would increase the likelihood of accessing the same interpretation in the target phase. The results confirmed this prediction, with participants being more likely to select a distributive scenario after distributive primes (in 68.33% of the cases) than after cumulative primes (58.17%) and baseline primes (61.15%), suggesting that activating a distributive interpretation during priming facilitated its subsequent reactivation. Cumulative primes did not show the same effect, i.e. participants did not select more often cumulative scenarios after cumulative primes than after baseline primes. The reason for this asymmetry is unclear, as both structures could theoretically trigger priming effects. However, the authors interpret this pattern as evidence that the distributivity operator was the primary driver of the priming effect.

Importantly, two additional control experiments ruled out visual priming as the source for these effects. In these experiments the target trials remained identical to those in the first experiment. However, in the priming phase of Experiment 2, participants saw the same pictures but were now presented with unambiguous sentences containing only one plural expression (e.g. *There are three circles*). In Experiment 3 the cumulative primes remained unchanged from the first experiment, but the distributive primes were modified such that the previously distributive picture (e.g. two squares each connected to three circles) was now paired with a new ambiguous sentence that was true in this scenario on a cumulative interpretation (e.g. *Two squares are connected to six circles* instead of *Two squares are connected to three circles*). This ensured that, while the visual stimuli remained identical, all prime trials were cumulative. Experiment

---

[5]On an *at least* interpretation of the numerals the sentences are compatible with both scenarios. Maldonado et al. (2017) argue against the possibility that participants accessed an *at least* interpretation citing the high accuracy rate in prime trials.

2 showed no priming effects, however, in Experiment 3 the manipulation led to an overall increase in cumulative choices, i.e. decrease in distributive choices, in target trials (33.27% in the cumulative condition compared to 41.59% in the baseline condition), suggesting that cumulative interpretations may also give rise to priming effects. Although the asymmetries in the data indicate that cumulative interpretations were weaker primes than distributive ones, the fact that specific interpretations of otherwise structurally identical sentences can be primed suggests that adults not only mentally represent the underlying meanings (or parts thereof), but also form distinct representations for cumulative and distributive readings.

There is also evidence suggesting that distributive interpretations of ambiguous plural sentences come with greater processing costs. In an eye-tracking experiment by Frazier et al. (1999) participants' eye movements were recorded while they read English sentences containing conjoined proper names in subject position and an ambiguous predicate followed either by *each* (19a), which enforces a distributive interpretation, or *together* (19b), which signals a collective (non-distributive) interpretation.

(19)   a.   *Sam and Maria / carried one suitcase each / at the airport.*

        b.   *Sam and Maria / carried one suitcase together / at the airport.*

Thus, the test sentences were ambiguous for participants up to the point where they encountered the disambiguating element *each* or *together*. In the regions of interest indicated by the slashes in (19), first pass reading times (i.e. the sum of fixations in a region between the time of first entering and first leaving the region), total reading times (i.e. the sum of all fixations in a region), and the pattern of regressions (i.e. regressive saccades out of a given region) were measured. In the region following *each/together*, longer first pass reading times (605 ms vs. 515 ms) and total reading times (768 ms vs. 592 ms) were found for sentences that contain *each* compared to sentences with *together*. Additionally, more regressions out of the third region into the second region were detected in sentences with *each* than in sentences with *together*. These findings suggest that participants initially favored collective interpretations, making it more difficult to process distributive interpretations upon encountering the disambiguating element, which not only provides further support for the view that two distinct representations underlie

16

those interpretations, but also that non-distributive interpretations are more basic, or serve as the default, while deriving distributive interpretations requires additional effort.

2.3   Acquiring the distributive/non-distributive contrast

While evidence suggests that adult speakers can access both distributive and non-distributive interpretations of ambiguous plural sentences, they also restrict their interpretation when disambiguating elements are present in the sentence. In contrast, children, even into school age, appear indifferent toward the distributive/non-distributive distinction. What is particularly striking is that preschoolers (and in some studies older children, e.g. Pagliarini et al., 2012) do not consistently assign a distributive meaning to quantifiers such as *each* or a non-distributive meaning to collective markers such as *together*. Instead, they tend to accept both interpretations for sentences containing these elements, even when adults perceive them as unambiguously enforcing a specific reading (e.g. Brooks and Braine, 1996; Musolino, 2009; Dotlačil, 2010; Pagliarini et al., 2012; Syrett and Musolino, 2013, 2016; Saponaro et al., 2024).[6]

This is suggested for instance by the results of a study by Pagliarini et al. (2012) which explored whether children and adults exhibit the same preferences regarding distributivity and non-distributivity of plural sentences, particularly whether the adult preference for the non-distributive reading of such sentences that some have observed (e.g. Brooks and Braine, 1996; Dotlačil, 2010) is already present in children across different age groups. In addition, sentences containing the quantifying expression *ciascun/a* ('each') were included to examine whether there is a correlation between children's interpretation of plural sentences and their acquisition of the semantics of 'each'. In a truth-value judgment task participants were presented with sentences similar to the examples below and asked to judge their truth or falsity given a picture showing either a collective scene, in which three characters performed an action on a single object (e.g. three girls building one snowman together), or a distributive scene, where each

---

[6]To our knowledge, there are no studies on the acquisition of cumulativity markers such as *between them*, which, however, might be expected given that the cumulative interpretation is cross-linguistically unmarked, i.e. languages do not seem to have obligatory cumulativity markers (cf. Haslinger et al., 2021).

character performed an action on a separate object (e.g. three girls each building their own snowman).

(20) a. *Le bambine costruiscono un pupazzo di neve.*

'The girls are building a snowman.'

b. *Ciascuna bambina costruisce un castello die sabbia.*

'Each girl is building a sand castle.'

When presented with collective scenes, adults overwhelmingly accepted sentences containing plural noun phrases (98%), but mostly rejected sentences with 'each' (9%). Conversely, when shown distributive scenes, they readily accepted quantified sentences (96%), but only accepted plural sentences half of the time (50%). Thus, adults seemed to access only a distributive interpretation for sentences with 'each', while accessing both interpretations for plural sentences, though they exhibited a lower acceptance rate for plurals in distributive scenarios than for sentences with 'each', which was interpreted as preference by the authors.

In contrast, younger children accepted both plural sentences and sentences containing 'each' in both collective and distributive scenarios. For instance, 5-year-olds accepted sentences with 'each' in collective scenarios in 92% of cases and plural sentences in 97%. Similarly, in distributive scenarios, they accepted sentences with 'each' in 100% of trials and plural sentences in 98%. With age acceptance rates for 'each' in collective scenarios gradually decreased, reaching an adult-like level by age 11. Plural sentences remained equally accepted in both distributive and collective scenarios until around age 10, when acceptance rates in distributive scenarios began to decline. Notably, even among 13-year-olds, acceptance rates for plural sentences in distributive scenarios remained higher than in adults.

These results confirm that in adult language 'each' enforces a distributive interpretation, whereas definite plurals are compatible with both collective and distributive readings. However, the findings also suggest that younger children do not yet fully grasp the semantic distinction between definite plurals and quantified phrases. Interestingly, there was a correlation between children's acceptance of plural sentences in distributive scenarios and their acceptance of quantified sen-

18

tences in collective scenarios. As acceptance of quantified sentences in collective contexts declined, so did acceptance of plural sentences in distributive contexts. This suggests a potential link between the acquisition of distributive markers and the preference for ambiguous plural sentences.[7]

The findings of Syrett and Musolino (2013) point in the same direction, showing that English-speaking preschool-aged children (ages 3-4) do not yet exhibit adult-like interpretations of plural sentences containing the collective marker *together* and the distributive marker *each*. This study investigated both the general availability of collective and distributive interpretations and whether speakers have a preference for one reading over the other. In the first experiment, adults and 3-to-5 year old children saw videos that either showed a collective event (e.g. two actors pushing one toy car together) or a distributive event (e.g. two actors pushing each one toy car) while listening to an ambiguous sentence such as (21).

(21)   *Two boys pushed a car.*

Children and adults accepted such sentences in collective scenes (3-year-olds: 96.7%; 4-year-olds: 98.3%; adults: 100%) and distributive scenes (3-year-olds: 91.7%; 4-year-olds: 98.3%; adults: 79.2%), suggesting that both groups considered either interpretation as a possible interpretation of ambiguous plural sentences.[8] In an subsequent preference task participants were presented with two still images that illustrated a distributive and a collective event and corresponded to the final scenes of the test videos used in Experiment 1 and were asked to select the one that went best with the presented sentence. Here children's and adults' responses diverged. Adults showed a strong preference for collective scenarios over distributive ones (88.9% vs. 11.1%), whereas children displayed a slight preference for distributive scenarios over collective ones (68.5% vs. 31.5%).

The results of two additional experiments (Experiments 4 & 5) suggest that disambiguating elements such as *each* and *together*, as in (22), had only a limited impact on children's judg-

---

[7]The interpretation of the data is not straightforward since test sentences were potentially scopally ambiguous. In Italian *un/una* is ambiguous between the singular indefinite ('a') and the numeral 'one'. Thus, it is not entirely clear whether the acceptance in collective scenarios was in fact du to a wide-scope interpretation of the indefinite.

[8]As in Pagliarini et al. (2012), it remains possible that collective interpretations were in fact wide-scope interpretations of the indefinite.

ments. However, *together* increased children's preference for collective scenarios when they were asked to choose between collective and distributive interpretations.

(22) a. *Two boys each pushed a car.*

   b. *Two boys pushed a car together.*

Specifically, while adults rarely accepted sentences containing *each* in collective scenes (31.9%), children overwhelmingly did (86.7%). Similarly, sentences containing *together* were rarely accepted by adults in distributive scenes (23.6%), whereas children accepted such sentences in most distributive trials (83.3%). On the other hand, in a preference-task children exhibited a clear preference for collective scenes over distributive ones when presented with sentences such as (22b) (71.3% vs. 28.7%). Taken together these findings suggest that preschoolers begin to develop intuitions about the meaning of the distributive marker *each* and the collective marker *together*, but they also indicate that children may not yet fully grasp the precise semantic contribution of these elements.

Similar results were obtained in a now classic study by Brooks and Braine (1996), which suggest that children are more indifferent about the precise meaning of the universal *each* compared to adults. Brooks and Braine (1996) examined the comprehension of the English universals *each* and *all* in 4- to 10-year-old children and adults. A series of three experiments investigated whether there is a preference for collective or distributive scenarios, whether preference varies depending on the lexical content of the universal and to what extent the concept of exhaustivity is deployed.

Preference was assessed through a picture selection task, whereby a sentence was read out to the participants while presenting two pictures, from which they could choose one. Among the sentence that were used in Experiment 1 were examples such as (23a) and (23b) which contained either the quantifier *each* or *all*.

(23) a. *All of the men are carrying a box.*

  b. *Each man is carrying a box.*

Pictures that were presented along with sentences containing *all* showed collective scenarios in which either three characters were performing an action on a single object, with two extra objects in the picture (e.g. two extra boxes), or one character was performing an action on three objects, with two extra characters in the picture (e.g. two extra men). Pictures that were presented with sentences containing *each* showed distributive scenarios in which either three characters individually performed the same action on three distinct objects, with two extra objects in the picture, or three characters individually performed the same action on three distinct objects, this time with two extra characters in the picture. Thus, the pictures varied in whether they depicted a collective or distributive scenario and whether there were additional characters or objects (i.e. exhaustivity of subjects or objects).

When presented with *all*-sentences, children's responses were almost adult-like, i.e. they preferred subject-exhaustive pictures for sentences with *all* such as (23a). For instance, already 5-year-olds preferred subject-exhaustive scenarios in 96.7% of cases compared to a 100% preference rate in the adult group.

Performance differed for sentences with *each*: Children apparently did not take into account the exhaustive meaning associated with *each* and frequently preferred the object-exhaustive picture with two extra actors in the scene for sentences like (23b). For instance, 5-year-olds preferred the subject-exhaustive picture in only 46.7% of trials, whereas in adults the rate was 100%. This again might indicate that younger children do not yet master the truth-conditions associated with *each*. Alternatively, as suggested by the authors, the results could reflect young children's ignorance toward the extra objects.

The second experiment investigated to what extent type of expression may influence the choice between distributive and collective scenarios. Test sentences contained either *each*, *all* or *three* as in the examples below.

21

(24) a. *All the men are building a boat.*

b. *Each man is building a boat.*

c. *Three men are building a boat.*

This time each sentence was presented along with pictures depicting either an exhaustive collective scenario, in which three characters were performing together an action on a single object, or an exhaustive distributive scenario, in which three characters were individually performing the same action on three distinct objects. Overall, collective scenarios were selected more often for sentences with *all* and *three*, whereas distributive scenarios were preferred for sentences with *each*. Adults preferred collective scenario for sentences with *all* such as (24a) in 83.3% of trials. Similarly, they preferred collective scenarios with *three* like the example in (24c) in 97.5% of cases. In contrast, adult preference rate for sentences with *each* was only 0.8%. The responses of the youngest age-group (4-year-olds) did not vary depending on the used expression, however, with age the choices became more adult-like and the number of collective choices for *all*-sentences and sentences with *three* increased, while the one for *each*-sentences decreased. For instance, while 5-year-olds selected the collective scenario for *all*-sentences in 54.4% of cases, 9-year-olds picked the collective scenario for such sentences in 72.2% of trials. As in Experiment 1, 5-year-olds did also not show a clear preference for distributive scenarios for sentences with *each*, with selecting collective scenarios for sentences like (24c) in 44.2% of cases.[9]

The third experiment examined the role of extra objects and involved sentences containing either *each*, *all* or a definite plural (25) which were presented simultaneously with three different types of pictures.

---

[9]Again the collective scenarios do not necessarily require a collective interpretation of *all*-sentences as they are also compatible with an inverse-scope reading, i.e. with a distributive wide-scope interpretation of the existential, illustrated below.

(i) *All of the men are carrying a box.*

(ii) a. **Surface scope:** $\forall x[\text{man}(x) \rightarrow \exists y[\text{box}(y) \wedge \text{carry}(x,y)]]$
'For every man there is a box that he carried.'
b. **Inverse scope:** $\exists y[\text{box}(y) \wedge \forall x[\text{man}(x) \rightarrow \text{carry}(x,y)]]$
'There is a box that was carried by every man.'

Moreover, both collective and distributive scenarios are in principle compatible with a surface scope interpretation (where it just happens to be the same box that each man carried).

(25)   a. *All the women are carrying a cake.*

       b. *Each woman is carrying a cake.*

       c. *The women are carrying the cakes.*

The pictures showed either a collective scenario, in which three characters performed an action on a single object together, with two extra objects in the picture (e.g. two extra cakes), a distributive scenario, in which three characters performed an action on three distinct objects, with two extra objects in the picture, or a cumulative scenario, in which three characters performed an action between them on five objects, with no extra objects in the picture (e.g. one woman carrying one cake, the other two women carrying two cakes each). As in the second experiment, the number of choices of collective scenarios for *all*-sentences and of distributive scenarios for *each*-sentences increased with age. For sentences with definite plurals the number of choices of cumulative scenarios increased with age. Preference for sentences with *all* was stable across age-groups with adults selecting collective scenarios in 58.3% of cases and 5-year-olds in 50%. For sentences with *each*, 5-year-olds decided for the distributive picture in 41.7% and adults in 97.2% of trials. Adults selected the cumulative scenario when presented with definite plural sentences in 97.2%, whereas this was only the case in 41.7% of trials in the group of 5-year-olds.

This study established that adults and children exhibit different preference patterns when interpreting plural and quantified sentences with *each* and *all*, indicating that they may assign different interpretations to these constructions. In particular, sentences containing *each* elicited different response patterns in terms of both exhaustivity and distributivity, suggesting that children struggled with both meaning components.[10]

Overall then previous research suggests that across languages children and adults diverge in

---

[10]The study by Brooks and Braine (1996) also involved a phase that was meant to test for general availability of distributive and non-distributive interpretations, however, their methodological choice may have influenced children's responses and thus we do not report the results here. In this phase, which always followed the main picture selection task, the experimenter picked four of the previously used test items and asked again which one of the two pictures went best with the sentence. After participants had decided for one, the experimenter asked whether the sentence could also go with the other picture. This means that participants were asked the same question about a subset of previously seen test items, plus a follow-up question whether they would also consider the other picture as a good match. Such follow-up questions are pragmatically quite problematic, e.g. in has been shown that they very easily lead children to revise their responses (cf. Bonawitz et al., 2020).

which interpretation they favor for different types of expressions that pick out multiple individuals, and moreover that before school-age, children have difficulties with the exhaustivity requirement of universals as well as the distributivity associated with certain quantifying expressions.

2.4   Targeting abstract levels of representation

The above reviewed studies, which used some version of the truth-value judgment task to assess children's comprehension of plural expressions and quantifiers as well as picture-selection tasks for measuring preference for distributive and non-distributive (cumulative or collective) interpretations, did not find evidence for children having acquired the meaning of distributive quantifiers such as *each* and its equivalents in other languages before school age. How can we make sense of the observed insensitivity toward the distributive/non-distributive contrast in young children? One way to interpret children's acceptance of a wide range of scenarios, even when presented with obligatorily distributive sentences, is to assume that they construct different, possibly simpler representations of sentences like *Each boy is building a snowman*, which are compatible with a broader set of situations, e.g. conveying a plural meaning. It is conceivable that this is because they do not yet have learned a distributive universal quantifier meaning and therefore rely on different representations of such sentences. Yet, the findings do not necessarily lead one to conclude that such meanings are not part of children's early grammar since for them distributive scenarios seem to be equally compatible with sentences containing a distributive universal as cumulative or collective scenarios. An alternative explanation would be that children do have access to a distributive quantifier meaning and derive distinct representations for distributive and non-distributive interpretations, but they may not yet be able to match the appropriate linguistic form with the corresponding meaning. However, the paradigms that were used to test children's competence provide limited insights into this issue.

To explore whether young children are able to distinguish and to represent distributive and

24

cumulative meanings of ambiguous plural sentences before they master the truth-conditions of overt distributive quantifiers, the priming paradigm seems particularly well-suited. Numerous studies suggest that properties at various levels of abstraction may be targeted by priming, including phonological properties, syntactic properties and semantic properties (see Branigan and Pickering, 2017 for an overview). Priming tasks consist of two consecutive phases. In the first phase, it is ensured that a particular representation is activated, after which, in the second phase, it is measured whether relevant features of the previously induced representation are easily (re-)activated. This allows drawing conclusions about the presence or absence of features as well as about their specifics. Relevant for us are comprehension studies that targeted structural ambiguities. Those usually attempt to enforce a particular representation of an ambiguous sentence during the priming phase by asking participants to choose between a correct item, which is congruent with the intended representation, and an incorrect item, thus guiding them towards the intended representation. Subsequently, it is tested whether participants tend to access the same type of representation for similar sentences when given the choice between a range of congruent items.

As mentioned earlier, Maldonado et al. (2017) found priming effects resulting from the cumulative/distributive ambiguity of plural sentences. Feiman and Snedeker (2016) used a similar design to study ambiguities of sentences that contain two quantifying expressions such as (26), which are assumed to be caused by structural differences at an abstract level (i.e. which do not correspond to different realizations of the sentences) and found that both of the interpretations in (26a) and (26b) may be primed in adults.

(26)   *Every hiker climbed a hill.*

    a.   **Surface scope:** For every hiker there is a hill that was climbed by that hiker.

    b.   **Inverse scope:** There is a hill such that every hiker climbed that hill.

A study by Havron et al. (2020) showed that even in children abstract structural properties may be primed. In their study a different type of structural ambiguity was targeted, namely syntactic attachment ambiguities of sentence such as (27). In one case the phrase *avec le pinceau* ('with the brush') attaches to the verb phrase, while in the other case it attaches to the

object noun phrase, leading to the two interpretations in (27a) and (27b). Both interpretations yielded priming effects in 5-6-year-old children.

(27)   *La fille chatouille le bébé avec le pinceau.*

‘The girl is tickling the baby with the brush.’

    a.   **High attachment:** The girl uses a brush to tickle the baby.

    b.   **Low attachment:** The girl is tickling the baby who is holding the brush.

This research may be taken as evidence that various structural properties are mentally represented and moreover implies that priming is a technique with the potential to reveal abstract representations, even in young participants. Importantly, in each of the aforementioned studies, different interpretations of superficially identical sentence structures were shown to produce distinct priming effects. We therefore adapted this method in a series of experiments aimed at detecting representational resources in preschool-aged children that traditional methods may have failed to reveal.

2.5   Plural ambiguities in children

The goal of our study was to investigate whether the cumulative/distributive contrast may also produce priming effects in 4-to-5-year-old children, i.e. at an age at which previous studies have demonstrated that they do not clearly distinguish between distributive and non-distributive interpretations and left open whether they can assign a distributive quantifier meaning to distributivity markers. We adapted the design developed by Maldonado and colleagues (2017), but conducted the study in German and simplified the task in terms of length and overall complexity. As in the original study, throughout the experiment sentences were used that contained two plurals modified either by the number word for ‘two’ or ‘three’, but did not involve any disambiguating elements (such as ‘each’, German *jed-*) and thus were ambiguous between the two readings, like the example in (28). We tested whether initially enforcing one of the interpretations would affect subsequent preference rate.

(28)   *Zwei Kühe ziehen drei Pferde.*

      'Two cows are pulling three horses.'

(29)   a.   **Cumulative interpretation:** Two cows between them are pulling three horses.

      b.   **Distributive interpretation:** Two cows each are pulling three horses.

Based on earlier findings, we were confident that children in that age range would know which cardinality is associated with the number words used in our study and would be able to access the *exact* interpretation of those (Wynn, 1992b; Musolino, 2004; Hurewitz et al., 2006; Sarnecka et al., 2007; Le Corre and Carey, 2007; Huang et al., 2013; Gelman et al., 2019; Rousselle and Vossius, 2021, a.o.). Control trials were included to ensure that this indeed was the case. We ran two experiments. In Experiment 1, we aimed to find out whether children have the ability to represent complex meanings that involve two plural expressions. We expected to find priming effects for the cumulative interpretation of such sentences if children can access the part-whole structure of pluralities and understand how the properties of individual members of the pluralities add up in relation to each other. Furthermore, if there are any developmental asymmetries, the ability to access cumulative interpretations can be expected to emerge earlier given its assumed simpler logical form.[11] As for the distributive interpretation, we hypothesized that if children's previously observed errors with distributive quantifiers are due to their inability to deploy a distributive quantifier meaning, we should not see priming effects at that age. Such a finding would be consistent with the Learning Hypothesis. If, on the other hand, we can detect priming effects for the distributive interpretation of such sentences, this would provide evidence that children are able to construct quantified representations and the indifference found in other studies could be attributed to a lack of knowledge of the rules that guide the mapping between linguistic forms and the corresponding meanings, supporting the Mapping Hypothesis. In Experiment 2, we used similar sentences but introduced greater variability in the stimuli to assess the generality of children's capacities.

---

[11]Though such a prediction does not necessarily follow from accounts which assume that the cumulative interpretation also involves the insertion of an operator (cf. Sternefeld, 1998; Beck and Sauerland, 2000).

2.6 Experiment 1

2.6.1 Methods

Participants

We tested 96 native German-speaking children who were randomly assigned to 3 experimental conditions: 32 participants to the Cumulative condition (range = 4;00 – 5;11 years, M = 5;02 years, SD = 0.60, 13 females), 32 participants to the Distributive condition (range = 4;00 – 5;09 years, M = 4;09 years, SD = 0.56, 11 females) and 32 participants to the Baseline condition (range = 4;00 – 5;11 years, M = 4;11 years, SD = 0.67, 16 females). Participants were tested in our lab and in the Natural History Museum (NHM) in Vienna. Parental consent was obtained before testing. Additional 11 children were tested, but excluded because they did not fulfill the inclusion criteria (see section 2.6.2 for our inclusion criteria). The study was approved by the ethical commission of CEU PU (PREBO, Psychological Research Board).

Design

Our study involved 3 conditions between-subjects (prime type: cumulative, distributive, baseline) and a forced-choice sentence-picture matching task in which participants were asked to match a sentence with one out of two pictures. Each test trial consisted of a target trial which was preceded either by a cumulative prime (cumulative condition), a distributive prime (distributive condition), or no prime (baseline condition). In target trials, participants were presented with an ambiguous plural sentence and two pictures of which one showed a cumulative and the other one a distributive scenario. Cumulative prime trials involved an ambiguous plural sentence presented along with two pictures of which one depicted a cumulative scenario and the other one a scenario which renders the sentence false on any interpretation. In distributive prime trials, an ambiguous plural sentence was accompanied by two pictures, one displaying a distributive scenario, the other one a false scenario (Figure 2.2). The dependent measure was

the rate of distributive choices in target trials. Filler trials followed the same structure, but involved unambiguous sentences presented together with two pictures of which one rendered the sentence true and the other one false (Figure 2.3).

Cumulative Prime

Cumulative False

"Two cows are pulling three horses."

Target

Distributive Cumulative

"Two rabbits are pulling two pigs."

Distributive Prime

False Distributive

"Two pigs are pulling two geese."

Target

Cumulative Distributive

"Two horses are pulling three cows."
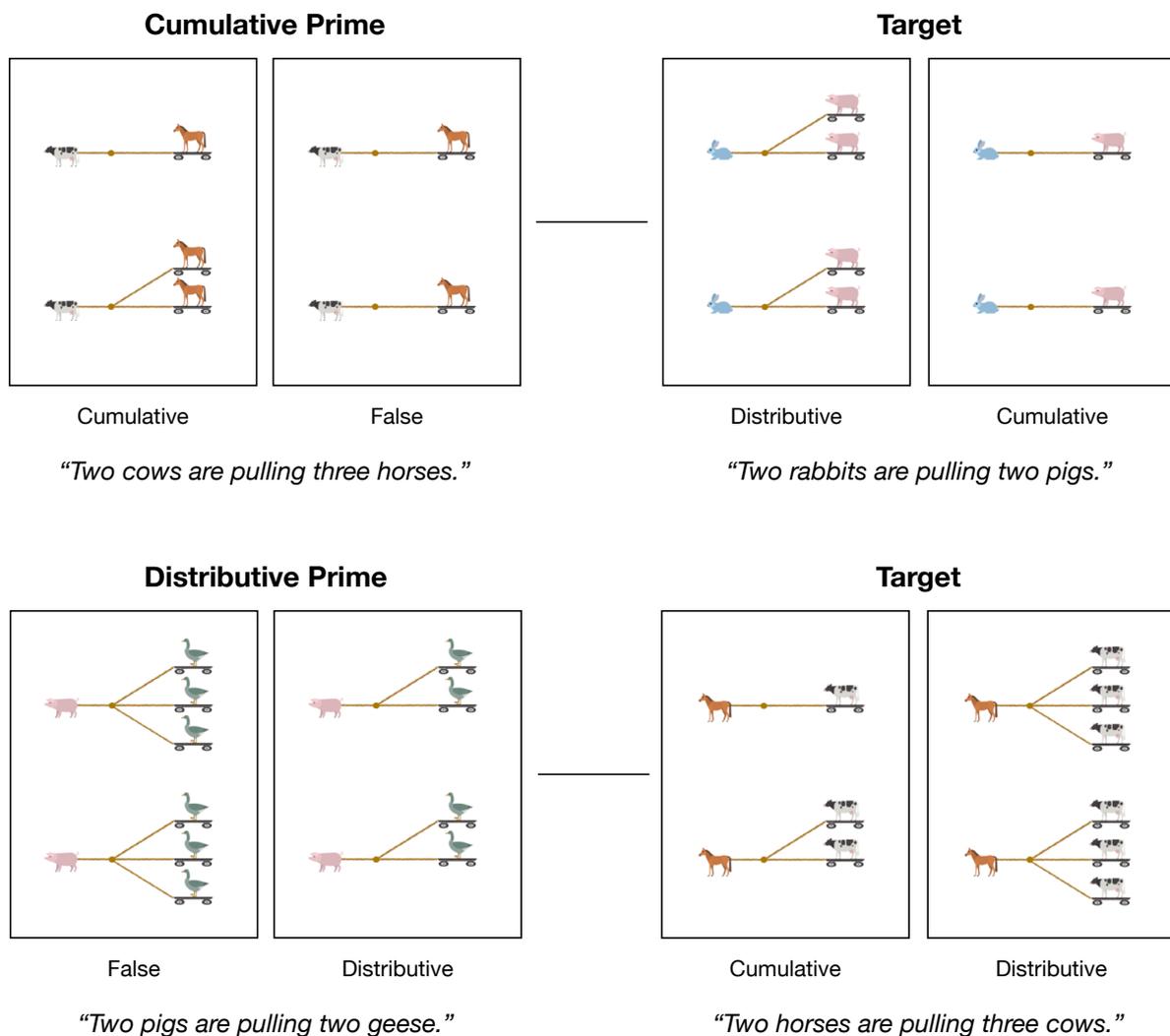
Figure 2.2: Test trial structure and stimuli exemplifying all possible combinations of numerals (Prime: 'three'/Target: 'two' for the cumulative condition and Prime: 'two'/Target: 'three' for the distributive condition) in Experiment 1

Materials

32 sentences were created using five different noun phrases denoting animals we expected to be familiar for the participants ('cow', 'pig', 'goose', 'horse' and 'rabbit'). Half of these sentences appeared in prime trials, the other half in target trials. Each sentence contained two

noun phrases in plural form modified by a numeral and, since lexical overlap has been shown to facilitate priming (cf. Mahowald et al., 2016), the verb 'to pull'. The relation expressed by the verb was meant to be relatively neutral in order to reduce impact of world-knowledge and lexical preferences. The numeral in subject position was held constant ('two'), while the object numeral was either 'two' or 'three'. Thus, prime and target sentences had one of the following structures:

(30) **Test sentences**

    a. Two *X*s pulled two *Y*s.

    b. Two *X*s pulled three *Y*s .

The syntactic position of the noun phrases and their combination within sentences were varied so that across all trials all nouns occurred in subject as well as in object position and no combination was repeated within prime and target trials. Each prime sentence was paired with a target sentence, yielding 16 test trials in total. Prime and subsequent target trials differed in noun phrases (i.e. no noun was repeated within one test trial) and object numerals (Prime: 'two'/Target: 'three' or Prime: 'three'/Target: 'two'). Additional 12 sentences were created using the same noun phrases and included as filler sentences to test children's attention and competence with number words as well as to ensure that the *exact* interpretation of the number words was accessed. These sentences contained one noun phrase modified by a numeral ('one', 'two' or 'three') and had one of the following structures:

(31) **Filler sentences**

    a. There is one *X*.[12]

    b. There are two *X*s.

    c. There are three *X*s.

To mimic the structure of the priming trials, filler sentences also came in pairs, resulting in 6 filler trials, whereby no noun phrase and no numeral was repeated within one trial. All

---

[12]The German equivalent to 'one' (*ein-*) is ambiguous between the numeral 'one' and the singular indefinite article 'a'. We instructed our speaker to stress *ein-* to disambiguate towards the '(exactly) one' reading, which, given the high accuracy rate in filler trials, evidently led participants to access the intended interpretation.

sentences were pre-recorded by a native-speaker of Austrian German (for a complete list of all test and filler sentences see Appendix A).

In addition to the sentences we created 120 pictures, of which 96 were used in test trials and 24 in filler trials. The 32 pictures that were used in prime trials in the cumulative condition depict cumulative scenarios showing two animals pulling either two or three animals from a different category in total and the 32 pictures that were presented in prime trials in the distributive condition display distributive scenarios in which two animals are each pulling two or three animals from a different category. For target trials the same 32 pictures showing cumulative and distributive scenarios were used in all conditions. Sentences in filler trials were presented along with pictures that exhibit one animal pulling either one, two or three animals from a different category. The position of the correct picture in prime trials and the distributive picture in target trials as well as the numerical combination between prime and target sentences was counterbalanced. This way we created 16 unique lists with 22 trials for each condition.
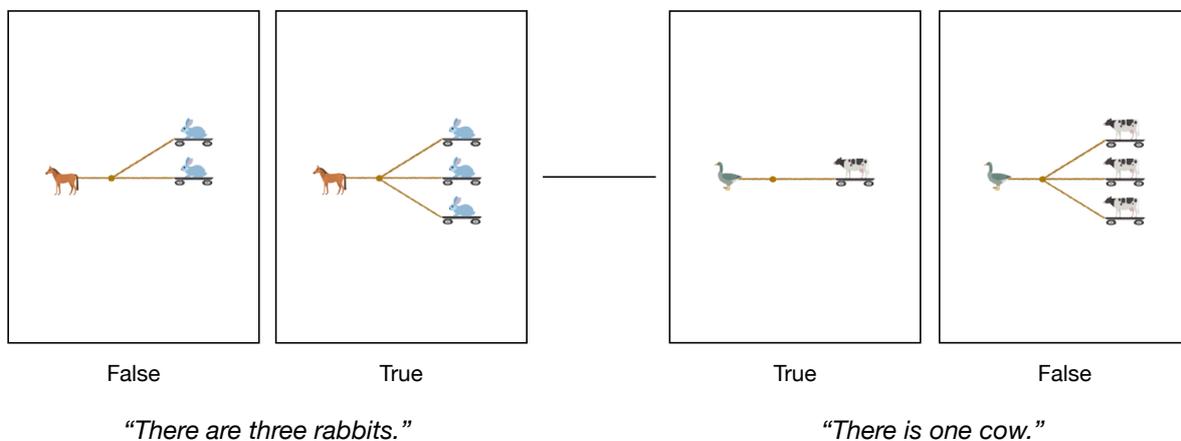

Figure 2.3: Filler trial structure and example stimuli in Experiment 1

Procedure

The experiment was created using PsychoPy 2022.2.4 (Peirce et al., 2019) and hosted by Pavlovia (Open Science Tools Limited, https://pavlovia.org). Stimuli were presented on an iPad with a 9.7-inch display. Participants were tested in a quiet testing room either in the lab or

in the museum in the presence of a caretaker. The task was introduced as a game with several rounds, in which the goal was to find the picture that matches best the sentence. Initially, a picture that depicted all main characters of the study (a cow, a pig, a goose, a horse and a rabbit) appeared on the screen and participants were asked to name each of these. No participant had difficulties at this phase, the items seemed to be highly familiar. Then children were explained that after listening to the sentence they could express their choice by touching the candidate on the screen. They were informed that they could not select both pictures or advance to the next round until they picked a picture. At the beginning of each trial, two pictures (picture size: 10.2 cm x 7.3 cm) positioned next to each other appeared on the screen. After a short period of 0.5 sec in which participants could visually explore the scene, a pre-recorded sentence was played while the two pictures remained on screen. The length of the sentences varied between 2 and 4 sec. Each trial took 4.5 sec, after which participants could make their choice. There was no time limit for making a choice and the pictures were on screen until participants responded. In the practice phase, 6 unambiguous sentences (such as 'There are three cows') were presented along with two pictures of which one showed a true scenario (e.g. a scenario with three cows) and one showed a false scenario (e.g. a scenario with two or one cow). Children were asked to make their choice and given feedback on their decision. During test phase, 16 test trials and 6 filler trials were presented in pseudo-random order and no feedback was provided. Participants were compensated with small toys or stickers independent of their performance.

2.6.2 Results

Only those participants were included in the final sample who had at least 75% of prime trials and at least 75% of filler trials correct (exclusion rate $\sim$ 10%). Participants in the final sample were highly accurate. Their overall accuracy rate in filler trials was at 98.33% (SE = 0.37), with 98.70% (SE = 0.54) correct choices in the cumulative group, 97.66% (SE = 0.67) in the distributive group and 98.96% (SE = 0.72) in the baseline group, which suggests that children understood the task and accessed the *exact* interpretation of the number words. In prime trials, accuracy was also high with children picking in 92.58% (SE = 0.92) of the cases the correct

32

picture. This was stable across conditions, with 93.55% (SE = 1.24) correct choices in the cumulative condition and 91.60% (SE = 1.36) correct choices in the distributive condition, indicating that both the cumulative and the distributive interpretation were equally accessible in the priming phase of the test conditions.
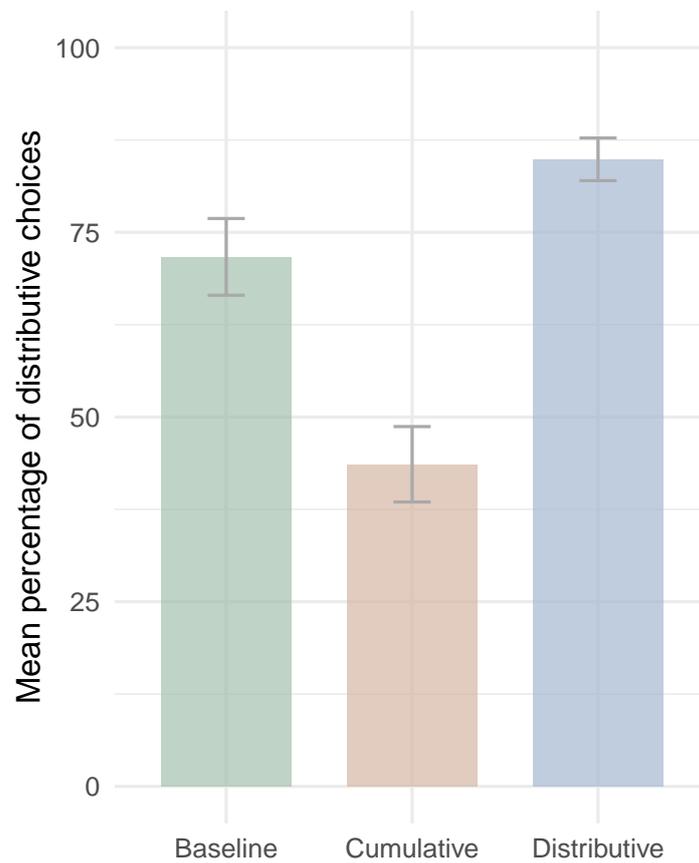
Figure 2.4: **Overall means** Mean percentage of distributive choices in target trials made by participants in each experimental condition in Experiment 1, with error bars representing the standard error.

For analysis only those target trials were considered that were preceded by prime trials in which the incorrect picture was rejected (cf. Raffray and Pickering, 2010). Less than 8% of trials were excluded from analysis. Figure 2.4 shows the mean percentage of distributive choices in target trials per condition. Children in the cumulative group selected in 43.60% (SE = 5.10) of valid target trials the distributive picture, while children in the distributive group did so in 84.88% (SE = 2.88) of the cases. In the baseline condition the distributive picture was chosen in 71.68% (SE = 5.19) of the trials.

The data was further analyzed with logistic mixed-effects models (Jaeger, 2008) in R (v4.0.5, R Core Team, 2021) using the lme4 package (Bates et al., 2015). Where relevant, the coefficients of the fixed effects in the mixed-effects models are reported as odds ratios (ORs) with 95% confidence intervals (CIs). For our main comparisons, we fit a model with Condition (3 levels: distributive, cumulative, baseline), Number (2 levels: 'two', 'three') and their interaction as fixed effects and included random intercepts for Participants and Items.[13] To assess the significance of the fixed effects we conducted model comparisons by removing the respective effect and used χ^2 likelihood ratio tests (following Maldonado et al., 2017, a.o.).

Main effects To test whether the rate of distributive choices in target trials was modulated by Condition, i.e. by the interpretation of the preceding prime sentence, we first compared the full model to a simpler one without Condition as predictor. We found a main effect of Condition: $\chi^2(2) = 34.65$, $p < .001$. There was also a significant effect of Number ($\chi^2(1) = 4.69$, $p = .03$) with children being more likely to select the distributive picture when the test sentence contained the numeral 'three' compared to 'two' across conditions (OR = 1.54, 95% CI [1.05, 2.25], $p = .026$) (see Figure 2.5).[14] The interaction between Condition and Number was not significant ($\chi^2(2) = 0.89$, $p = .642$). To determine whether the rate of distributive choices in the distributive condition differed significantly from choice rate in the cumulative condition, we compared a model that included only those two conditions to a simpler model. The comparison showed a significant effect of Condition ($\chi^2(1) = 35.19$, $p < .001$). Cumulative and baseline condition also differed significantly from each other ($\chi^2(1) = 14.7$, $p < .001$). When looking at a model that only included the distributive and the baseline condition, the effect of Condition was significant ($\chi^2(1) = 3.91$, $p = .048$).

Age effects We additionally explored the possibility that Age (centered, in months) contributed significantly to the observed effects, however, this hypothesis was not borne out ($\chi^2(1) = 0.08$,

---

[13]We did not include a random slope for Item within Participants since there was only one observation per each Item unit per participant. The initial model included a random slope for Number within Participants and Number within Items, but failed to converge. Our final model included the maximal random effects structure appropriate for our design that converged (Barr et al., 2013).

[14]When looking separately at trials with the number word 'two' and at trials with the number word 'three', the effect of Condition was still significant in each subgroup ($\chi^2(2) = 31.09$, $p < .001$ for 'two' and $\chi^2(2) = 27.47$, $p < .001$ for 'three').
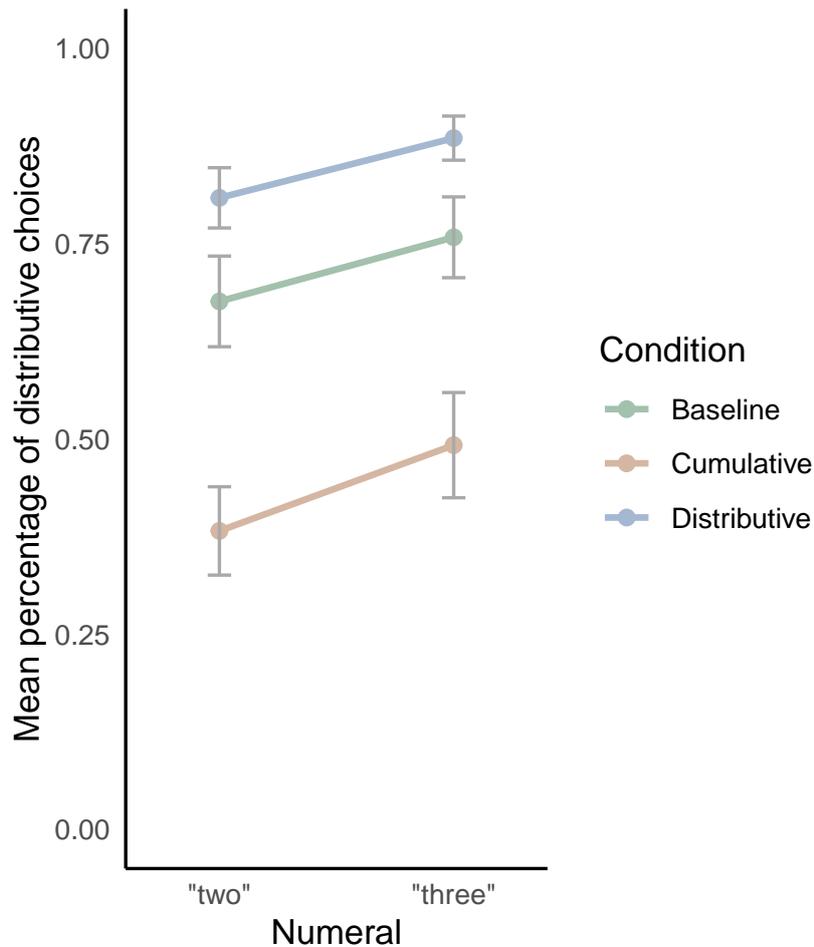
Figure 2.5: **Means by numeral** Mean percentage of distributive choices in target trials by numeral ('two' vs. 'three') across experimental conditions in Experiment 1. Points represent the mean response per condition and numeral and error bars indicate the standard error.

$p = .776$). There was also no interaction between Condition and Age ($\chi^2(2) = 4.21$, $p = .122$) (see Figure 2.6).

Comparison against chance To assess whether responses in each experimental condition differed from chance, we fit a simpler model excluding Number as a fixed effect. For each condition, we specified a model with that condition as the reference level and tested whether the intercept differed from a chance level of 50% (corresponding to an odds ratio of 1). The analysis revealed that in the cumulative condition, responses did not significantly differ from chance (OR = 0.64, 95% CI $[0.32, 1.28]$, $p = .209$). In contrast, the choice rate in the distributive condition was significantly above chance (OR = 12.91, 95% CI $[5.95, 28.03]$, $p < .001$). Similarly, participants in the baseline condition were also significantly more likely to pick the distributive

picture than expected by chance (OR = 4.59, 95% CI [2.22, 9.49], $p < .001$).
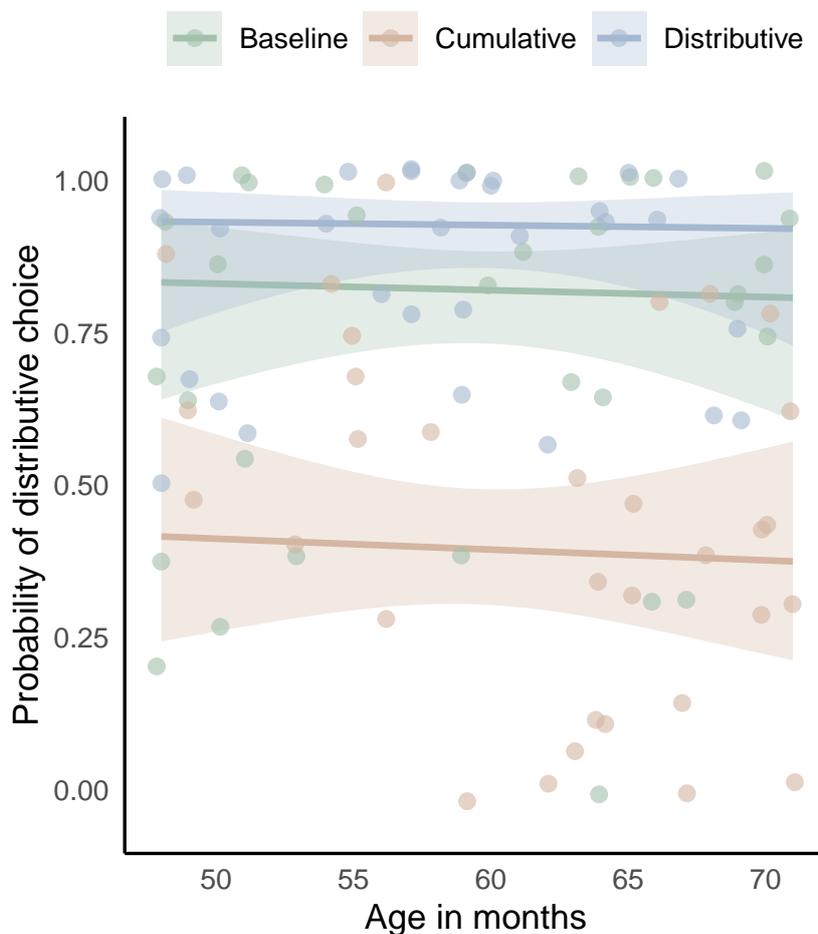
Figure 2.6: **Predicted probabilities and participant means** Probability of a distributive choice in target trials by age across experimental conditions in Experiment 1. Smoothed lines with confidence intervals depict model-predicted probabilities, while dots represent individual participant means for each condition and age.

2.6.3   Discussion

Overall our findings show that both the cumulative and the distributive interpretation of ambiguous plural sentences, which do not contain any overt markers that enforce the respective reading, may lead to priming in preschool-aged children. As illustrated in Figure 2.4, the proportion of distributive choices was higher in the distributive condition (84.88%) than in the cumulative condition (43.60%) and also higher than the baseline preference (71.68%). Children who were presented with distributive primes during the priming phase were more likely

to select the distributive picture in target trials compared to children who were presented with cumulative primes. Children in the distributive group were also more likely to choose the distributive picture than children who saw no primes at all. At the same time, the group of children who were primed with the cumulative interpretation showed a decreased likelihood of picking the distributive picture compared to the baseline group. This shows that prime type had a significant effect on participants' preference and suggests that the interpretation that children accessed during the priming phase was more readily available in subsequent target trials.

The fact that compared to the baseline condition the increase in distributive choices in the distributive condition was smaller than the decrease in the cumulative condition could be due to an already high baseline preference for the distributive interpretation such that the level of activation could not be much further increased in the former case. This would also be in line with previous findings which showed that in general less preferred or less frequent constructions may lead to stronger priming effects, the so-called *inverse-preference effect* (Bock, 1986; Hartsuiker and Westenberg, 2000; Bott and Chemla, 2016). However, because of the between-subjects design it is difficult to evaluate our data in this respect as we do not know to what extent the effects in our test conditions were correlated with individual baseline preferences for one of the interpretations.[15]

The asymmetry that we found between the effects for the two interpretations is in principle compatible with a structural containment relation, i.e. the assumption that the distributive interpretation is derived from the cumulative interpretation through an additional operator (cf. Link, 1987), whereby the cumulative structure would prime the cumulative reading, while the more complex distributive structure would activate both the distributive interpretation and the basic cumulative interpretation, resulting in a smaller difference between distributive and baseline condition compared to the difference between cumulative and baseline condition. The results

---

[15]Though Maldonado et al. (2017) did not find a stronger priming effect for the dispreferred, in their case cumulative, reading, a similar study by Jiang et al. (2022), which involved a blocked within-subjects design for a more neutral baseline assessment, suggests that at the level of individual speakers stronger priming effects may arise for the initially less preferred interpretation. Overall previous evidence is inconclusive as to which interpretation of ambiguous plural sentences is generally favored by speakers. While many studies have found a preference for the cumulative interpretation (e.g. Dotlačil, 2010; Syrett and Musolino, 2013), some findings suggest that the distributive interpretation is preferred (Maldonado et al., 2017; Jiang et al., 2022).

are not compatible with the idea that the priming effect is solely driven by the presence of a distributivity operator, as, in contrast to the findings in (Maldonado et al., 2017), the effect observed in our study shows the reverse pattern. If the distributivity operator were the only source of priming, we would expect that choices in the cumulative and in the baseline condition, both characterized by the absence of the operator, would pattern alike and only distributive primes would lead to an increase in distributive choices in target trials. This, however, is not what we see, rather the data suggests that the cumulative structure may facilitate the re-activation of the cumulative interpretation just as the distributive structure may facilitate the re-activation of the distributive interpretation.[16] On the other hand, it is conceivable that the high baseline preference for the distributive interpretation may prevent us to a certain extent to estimate the actual degree of distributive priming. One possible explanation for the high proportion of distributive choices in the baseline condition could be a lexical preference of the verb 'to pull' for distributive interpretations. This possibility cannot be ruled out here. Another explanation could be that, given the visual features of some of the cumulative scenarios used in the study, the verification of the cumulative interpretation was somewhat more difficult than the verification of the distributive interpretation. In particular, children's responses might have been guided in parts by a general preference for symmetry which made the scenarios with symmetric distributions of patients more accessible than the ones with asymmetric distributions. In fact this is partly supported by our results, which show that the probability of selecting the distributive picture was higher in trials that involved sentences with the numeral 'three', where children could choose between a symmetric, distributive scenario and an asymmetric, cumulative scenario, in comparison to trials that involved the numeral 'two', in which both scenarios had an equal distribution of agents. But since this was true across conditions, this does not entirely explain the high rate of distributive responses in the baseline condition.

Nonetheless, the overall priming effect cannot be accounted for by a shallow verification strategy according to which participants based their choice on some superficial feature(s) of the visual stimuli. This becomes clear when looking at the similarities and differences of the test

---

[16]At this point it is important to note that our data is not fully comparable to the adult data collected by Maldonado and her colleagues (2017) since their study involved a within-subjects design.

pictures in prime and target trials in each condition (Figure 2.2). For instance, in the cumulative condition the prime sentence 'Two cows are pulling three horses' is rendered false by a scenario in which there are two cows of which each is pulling one horse. In the subsequent target trial, the sentence 'Two rabbits are pulling two pigs' on its cumulative interpretation is rendered true through a scenario in which there are two rabbits of which each is pulling a pig, i.e. a scenario which is characterized by exactly the same distribution of agents and patients as the false scenario in the preceding prime trial. Similarly, in the distributive condition the scenario that renders the prime sentence 'Two pigs are pulling two geese' false on its distributive interpretation, corresponds visually to the "correct" distributive scenario in the subsequent target trial. Thus, if children's responses would rely on some shallow matching strategy of the scenes, we would not expect to find any of the attested effects.

Finally, concerning our main question of whether children represent the cumulative/distributive contrast at an abstract level before they have fully mastered the truth-conditions of overt quantifying expressions like 'each', our findings suggest that this is indeed the case. First, the results show that choice rate in target trials is significantly modulated by the manipulations in the priming phase, suggesting that accessing an interpretation facilitated its re-activation in the subsequent phase of the experiment and that the direction of the effect was different for the cumulative and for the distributive group. Second, the effects can only be explained by participants accessing an abstract level of representation since at the surface level the presented sentences do not differ from each other, i.e. superficially the used test sentences are structurally identical. Therefore, the pattern of responses leads us to the conclusion that children indeed distinguish between cumulative and distributive interpretations at an abstract level even before mastery of the respective overt markers.

One might wonder why children in our task performed better than children of the same age in previous studies. We speculate that this is because in the used picture-selection task, particularly during the priming phase, the relevant interpretation of the ambiguous sentences was made salient to participants and thus, compared to the affordances of previously used truth-value judgment tasks to investigate children's comprehension of distributivity (e.g. Brooks and

Braine, 1996; Musolino, 2009; Pagliarini et al., 2012; Syrett and Musolino, 2013), they had to spend less effort in searching the logical space for possible interpretations. In consequence, the findings can also be taken as further support for the Mapping Hypothesis.

2.7   Experiment 2

In Experiment 1, it was established that children may derive both cumulative and distributive interpretations of ambiguous plural sentences and that, due to the lack of overt disambiguating elements and the variation in noun phrases and object numerals in our test sentences, the priming effects can only be accounted for when we assume that children formed abstract representations of those interpretations. However, since all test sentences in Experiment 1 involved the verb 'to pull', the question remains whether the effects would generalize to other verbs and whether they would hold across different verbs. Experiment 2 was designed to probe the level of abstraction and to more precisely identify the features underlying the effects observed in Experiment 1. To achieve this, we once again adapted the same paradigm, but this time introduced greater variation in the lexical items: While target sentences continued to include the verb 'to pull' (28), prime sentences featured the verb 'to push' as in (32).

(32)   *Zwei Kühe schieben drei Pferde.*

'Two cows are pushing three horses.'

The rationale was that if we find that priming occurs independent of the repetition of the verb, this would suggest that the effects in the first experiment were based only on the formal properties of distributive and cumulative interpretations. But if priming presupposes lexical overlap, we may conclude that priming in the first experiment arose because children formed cumulative and distributive representations of sentences that involve the relation 'to pull', but that this was not indicative of their general capacity to form such representations and potentially related to their conceptual knowledge associated with the specific lexical item.

2.7.1 Methods

This experiment was preregistered (https://aspredicted.org/5mdw-nsb6.pdf).

Participants

For Experiment 2, we tested a new group of 64 native German-speaking children. Participants were randomly assigned to 2 conditions: 32 participants to the Cumulative_dif condition (range = 4;00 – 5;09 years, *M* = 5;00 years, *SD* = 0.47, 14 females) and 32 participants to the Distributive_dif condition (range = 4;00 – 5;11 years, *M* = 5;01 years, *SD* = 0.62, 16 females). Test sessions took place in our lab and in the Natural History Museum (NHM) in Vienna and parental consent was obtained before testing. Additional 3 children were tested but had to be excluded because they did not meet the inclusion criteria (see section 2.7.2 for details). Like Experiment 1, this study was approved by the ethical commission of CEU PU (PREBO, Psychological Research Board).

Design and Materials

The study design was similar to the one we adapted for Experiment 1, except that this time prime sentences and target sentences involved different verbs. In both conditions, prime sentences contained the verb 'to push' and the corresponding pictures depicted a pushing action, while target sentences contained the verb 'to pull' with the corresponding pictures showing a pulling action (Figure 2.7). Thus, prime and target sentences had the following structure:

(33) **Prime sentences**

    a. Two *X*s pushed two *Y*s.

    b. Two *X*s pushed three *Y*s .

(34) **Target sentences**

    a. Two *X*s pulled two *Y*s.

b. Two *X*s pulled three *Y*s .

Similarly, filler trials were changed such that half of the filler pictures displayed a pushing
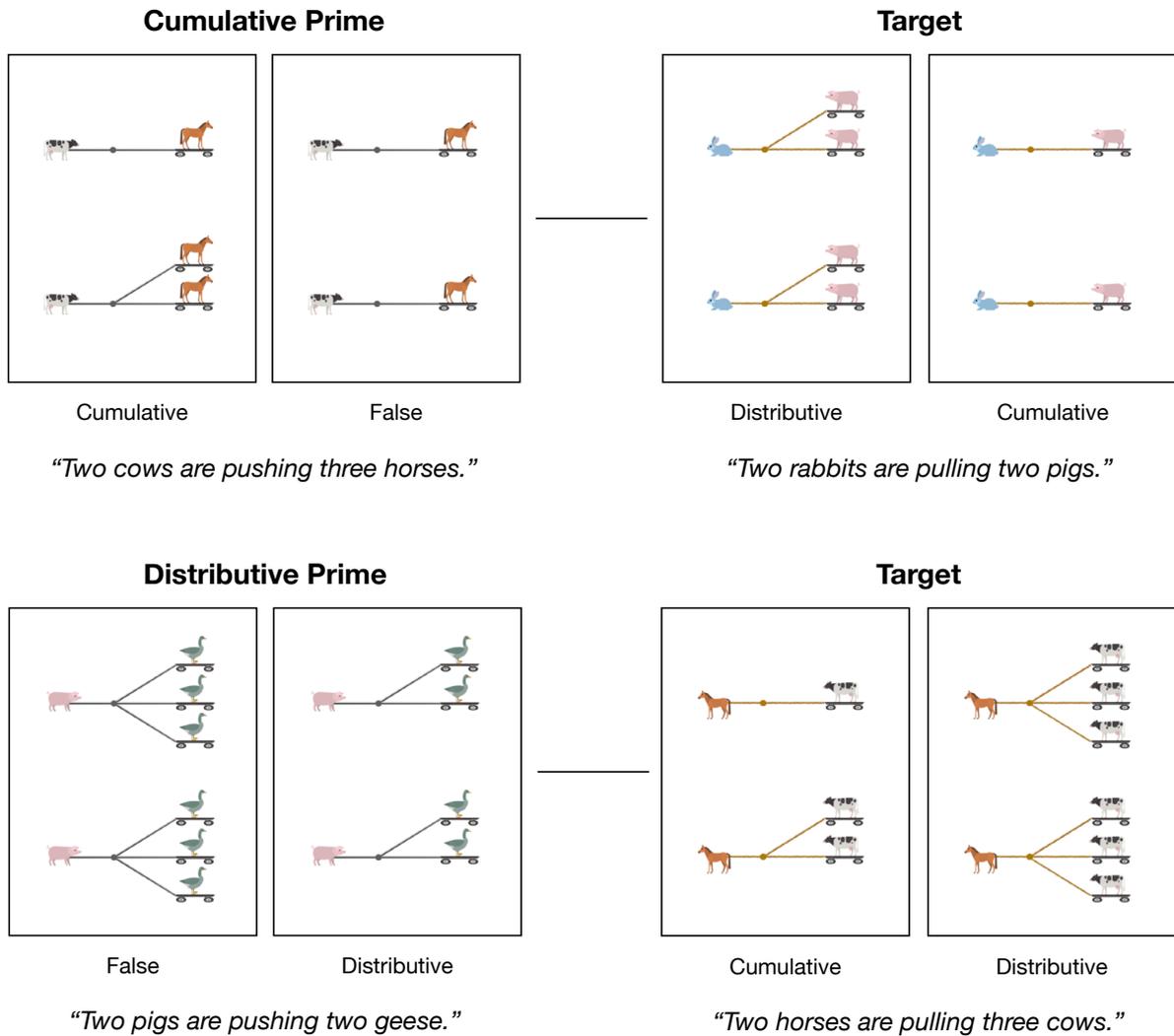
Figure 2.7: Test trial structure and stimuli in Experiment 2

action. A total of 32 sentences were used, of which one half involved 'to push' and the other half 'to pull' combined with two plural nouns modified by a numeral. Each prime sentence contained the verb 'to push' and was paired with a target sentence containing the verb 'to pull', resulting in 16 test trials (see Appendix A for a list of all sentences that were used in test trials in Exp. 2). Another 12 sentences were presented in filler trials. These were identical to the filler sentences used in Experiment 1. We created 96 pictures to be presented in test trials and 24 additional pictures to be presented in filler trials. 32 of these pictures showed a cumulative

scenario in which two animals are pushing either two or three animals from a different category in total and were used in prime trials in the cumulative_dif condition. The 32 pictures that were presented in prime trials in the distributive_dif condition depicted a distributive scenario in which two animals are each pushing either two or three animals from a different category. Target trials in both conditions involved 32 pictures showing cumulative and distributive scenarios in which two animals are pulling two or three other animals. As in Experiment 1, the dependent measure was rate of distributive choices in target trials. In filler trials, the first sentence always was accompanied by a picture displaying one animal pushing either one, two or three other animals, while the second sentence was accompanied by a picture that showed one animal pulling either one, two or three animals (Figure 2.8).
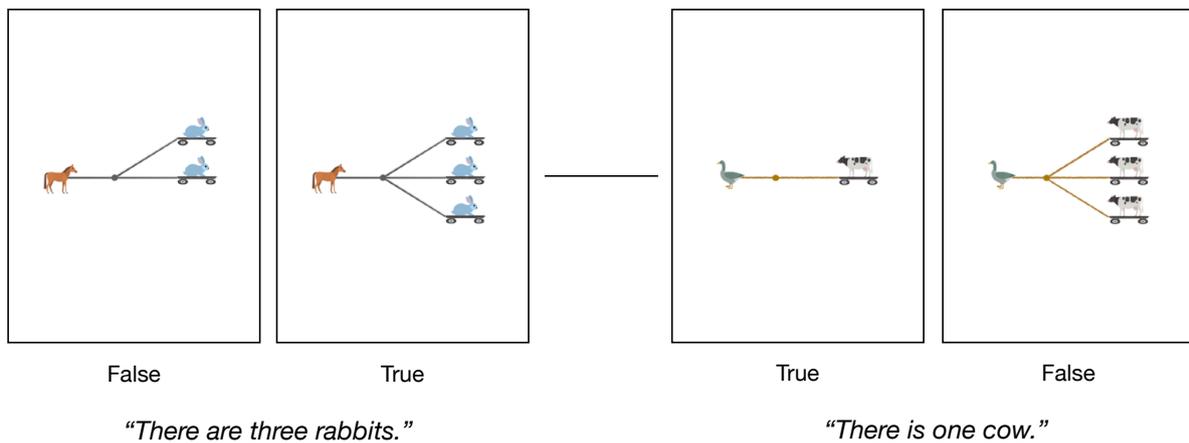
Figure 2.8: Filler trial structure and example stimuli in Experiment 2

Procedure

The procedure was identical to Experiment 1.

2.7.2   Results

Participants who did not choose the correct picture in at least 75% of prime and filler trials were excluded (exclusion rate $\sim 5\%$). Accuracy in filler trials was again very high, whereby children selected the correct picture in 98.70% (SE = 0.54) of the cases in the cumulative_dif

43

condition and in 97.66 (SE = 0.77) of the cases in the distributive_dif condition. Children across conditions were also highly accurate in prime trials. The overall accuracy rate in prime trials was at 93.07% (SE = 0.84), with participants picking the correct picture in 92.58% (SE = 1.27) of the trials in the cumulative_dif condition and in 93.55% (SE = 1.10) of the trials in the distributive_dif condition. Together this not only suggests that children were competent with the meaning of the number words, but also that they were able to distinguish impossible from possible interpretations of the ambiguous test sentences.
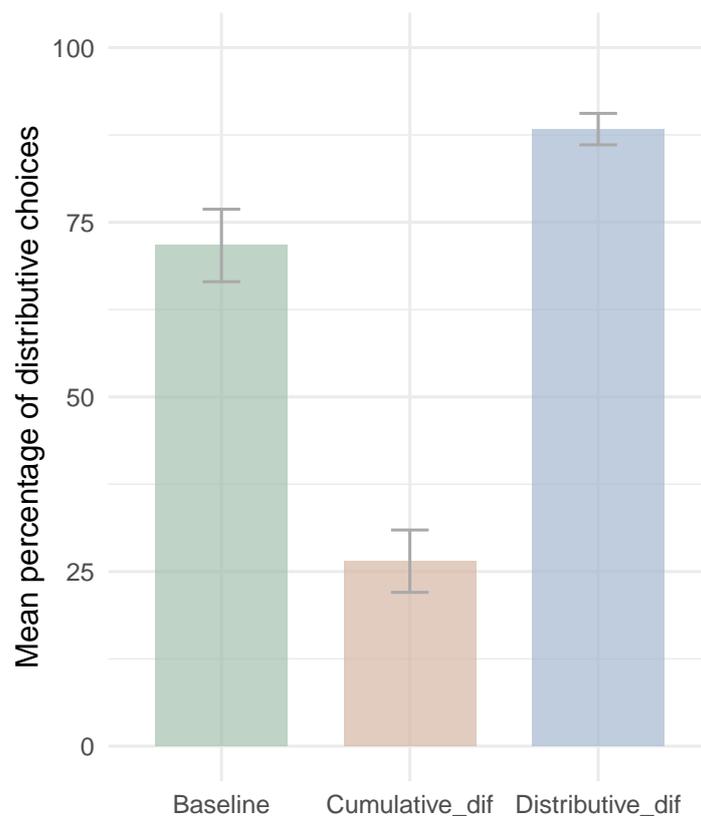
Figure 2.9: **Overall means in Exp. 2** Mean percentage of distributive choices in target trials made by participants in each experimental condition in Experiment 2, with error bars representing the standard error.

The data were analyzed as in Experiment 1. We considered only those target trials that were preceded by correct choices in the corresponding prime trials (exclusion rate $\sim$ 7%) and included the baseline condition from Experiment 1. The proportion of distributive choices per condition is illustrated in Figure 2.9. In the cumulative_dif condition, the mean percentage of distributive choices was 26.49% (SE = 4.46), compared to 88.33% (SE = 2.25) in the distribu-

tive_dif condition, and to 71.68% (SE = 5.19) in the baseline condition from Experiment 1. For the main comparisons, we fit a model that included Condition (3 levels: distributive_dif, cumulative_dif, baseline), Number (2 levels: 'two', 'three') and their interaction as fixed effects and random intercepts for Participants and Items. The significance of the fixed effects was tested by removing the relevant fixed effect and comparing the more complex model to the simpler model with χ^2 likelihood ratio tests.

Main effects Our analysis revealed a main effect of Condition ($\chi^2(2) = 66.15$, $p < .001$) and of Number ($\chi^2(1) = 12.30$, $p < .001$) with the Number 'three' being associated with an increased probability of a distributive choice (OR = 2.06, 95% CI $[1.48, 2.89]$, $p < .001$). When comparing models that only included the distributive_dif and the cumulative_dif condition, we also found a significant difference between the cumulative_dif condition and the distributive_dif condition ($\chi^2(1) = 63.56$, $p < .001$). Moreover, the cumulative_dif condition as well as the distributive_dif condition differed significantly from baseline ($\chi^2(1) = 31.84$, $p < .001$ and $\chi^2(1) = 6.37$, $p = .012$). This time there was an interaction between Condition and Number ($\chi^2(2) = 14.49$, $p < .001$) in that the likelihood of a distributive choice was further increased in trials with the numeral 'three' compared to trials that involved the numeral 'two' in the distributive_dif condition (OR = 3.39, 95% CI $[1.39, 8.23]$, $p = .007$), while this was not the case in the cumulative_dif condition (OR = 0.66, 95% CI $[0.33, 1.31]$, $p = .236$) (see Figure 2.10). Yet, the effect of Condition was still significant when looking posthoc separately at trials with 'two' ($\chi^2(2) = 50.31$, $p < .001$) and trials with 'three' ($\chi^2(2) = 74.43$, $p < .001$).

Age effects We additionally assessed whether Age (centered, in months) had an effect on the likelihood of distributive choices, which was again not confirmed ($\chi^2(1) = 0$, $p = .998$). There was also no interaction between Age and Condition ($\chi^2(2) = 4.48$, $p = .107$) (Figure 2.11).

Exp. 1 vs. Exp. 2 To see whether the repetition of the verb in prime and target trials in Exp. 1 affected distributive choices, we fit a model with Condition (2 levels: distributive, cumulative) and Experiment (2 levels: Experiment 1, Experiment 2) and their interaction as fixed effects. Though we detected an interaction between Condition and Experiment which was significant ($\chi^2(1) = 4.14$, $p = .042$) and showed that the difference between the distributive_dif and the

45

cumulative_dif condition in Exp. 2 was more pronounced than the one between the distributive and cumulative condition in Exp. 1 (OR = 3.53, 95% CI [1.06, 11.78], p = .04), our analysis did not reveal a main effect of Experiment on target responses ($\chi^2(1)$ = 1.95, p = .163), suggesting that the effect of Condition on distributive responses was largely consistent between Experiment 1 and Experiment 2.
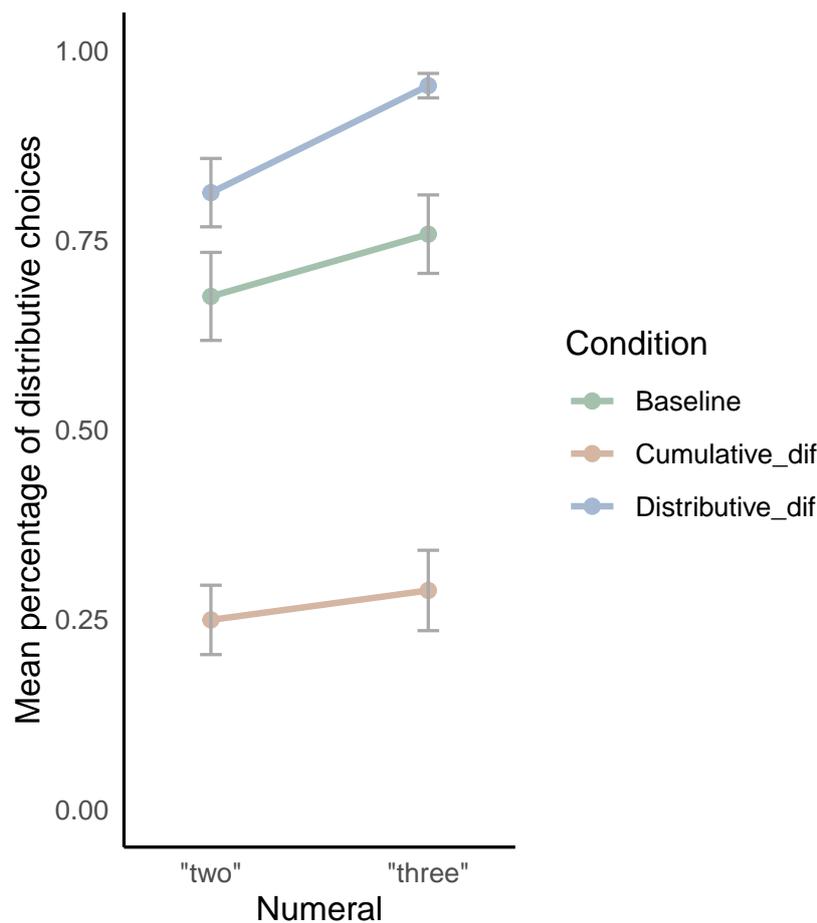
Figure 2.10: **Means by numeral in Exp. 2** Mean percentage of distributive choices in target trials by numeral ('two' vs. 'three') across experimental conditions in Experiment 2. Points represent the mean response per condition and numeral, with error bars indicating the standard error.

Comparison against chance Testing against chance again was performed with a simpler model that excluded Number as fixed effect. For each condition, we fit a model with that condition as the reference level and tested whether the intercept differed from a chance level of 50% (corresponding to an odds ratio of 1). The analysis showed that in the cumulative_dif condition the likelihood of a distributive choice was significantly below chance level (OR = 0.21, 95% CI [0.11, 0.41], p < .001), whereas in the distributive_dif condition it was significantly above

chance (OR = 15.52, 95% CI $[7.41, 32.53]$, $p < .001$).
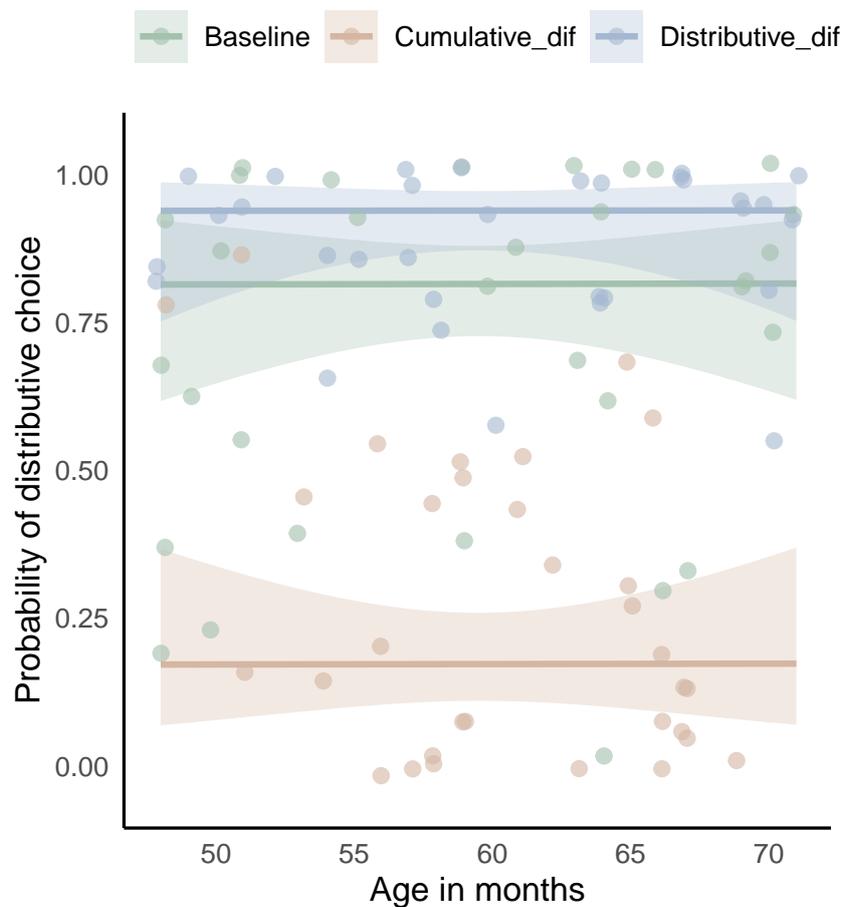
Figure 2.11: **Predicted probabilities and participant means in Exp. 2** Probability of a distributive choice in target trials by age across experimental conditions in Experiment 2. Smoothed lines with confidence intervals show model-predicted probabilities, dots represent individual participant means for each condition and age.

2.7.3 Discussion

By and large the pattern of responses in Experiment 2, in which prime and target sentences contained different verbs, closely resembles the pattern observed in Experiment 1, in which prime and target sentences shared the same verb. In Experiment 2 we found that even in the absence of lexical overlap in the verb phrase both the cumulative and the distributive interpretation of ambiguous plural sentences can lead to priming. The proportion of distributive choices in the distributive_dif condition (88.33%) was higher than in the cumulative_dif condition (26.49%) and in the baseline condition (71.68%). Children exposed to distributive primes with the verb

'to push' were more likely to select the distributive picture in subsequent target trials with the verb 'to pull' than those presented with cumulative primes or no prime at all. Conversely, in the cumulative_dif condition, children were less likely to choose the distributive picture in target trials with the verb 'to pull' after cumulative primes with 'to push' compared to those in the baseline condition, who were not exposed to any primes.

Crucially, sentences in prime and target trials in both the distributive_dif and the cumulative_dif condition differed not only in the verb used but also in the subject and object nouns as well as the object numeral. The only element that was repeated between prime and target was the subject numeral, suggesting that the priming effects did not rely on lexical knowledge. More specifically, what this seems to indicate is that the effects were solely driven by the formal properties that characterize the cumulative and the distributive interpretation of the sentences and not by any lexical information associated with the noun phrases and verbs or specific numerical information conveyed by the object numerals. This might be taken as evidence for the existence of a level of representation that is bare of lexical information and only encodes the logical properties of sentences. Alternatively, the results can be interpreted as demonstrating that priming may selectively target certain logical properties of a possibly richer representation. In any case the findings imply that children have the general capacity to form representations that encompass the abstract features of cumulative and distributive interpretations of ambiguous plural sentences.

One open question is why the priming effects were amplified in Experiment 2 compared to Experiment 1. Though many previous studies have found larger priming effects for prime-target pairs in which lexical items, especially verbs, were repeated (the so-called *lexical boost*) (Pickering and Branigan, 1998; Cleland and Pickering, 2006, a.o.), some research suggests that such effects last longer when the verb is not repeated between prime and target (e.g. Bock and Griffin, 2000). Based on these findings we cannot exclude the possibility that the enhanced priming effects in Experiment 2 were at least partly due to an increased availability of the respective interpretation throughout the experiment (see Figure A.1 and Figure A.2 in Appendix A for means by trial in Exp. 1 and Exp. 2). Apart from that, it seems also plausible that the

less repetitive nature of the task in Experiment 2 led participants to maintain a higher level of attention until the end of the task, as tentatively supported by the lower exclusion rate ($\sim 10\%$ in Exp. 1 vs. $\sim 5\%$ in Exp. 2), and that this boosted their overall performance and the priming effect.

Interestingly, neither in Experiment 1 nor in Experiment 2 did the age of our participants affect their performance. In particular, our data indicate that no developmental changes occurring between the ages of 4 and 6 years, whether related to general and/or linguistic abilities, influence children's comprehension of cumulative and distributive readings of ambiguous plural sentences. We interpret this as further evidence for the independence of the underlying mental representations.

2.8 General discussion

We began by contrasting two approaches to conceptual development and the mechanisms underlying the acquisition of functional vocabulary. According to Learning Theories, infants possess a small set of primarily domain-specific conceptual primitives that enable them to reason about their surrounding environment. Over time, exposure to external symbolic systems, such as language, facilitates the gradual development of richer and more powerful conceptual representations. Consequently, acquiring the meanings of logical expressions involves not only learning the relevant linguistic forms but also constructing the corresponding abstract meanings. According to Mapping Theories, infants' logical representations are already rich and the acquisition process therefore comprises identifying the relevant linguistic forms and mapping them onto antecedently existing meanings.

Since earlier research gives reason to suspect that preschoolers do not interpret distributivity markers in terms of a distributive universal quantifier, we asked whether their errors are due to them not having the representational capacities to derive distributive meanings or whether these may be explained by difficulties in the mapping between form and meaning. Priming

distributive and cumulative interpretations of plural sentences which do not contain explicit disambiguating markers allowed us to attribute any effect to children's abstract representations of such sentences.

Overall our findings speak in favor of the Mapping Hypothesis as they illustrate how powerful already preschoolers' representational system is. The results suggest that children from the age of 4 are not only able to represent a cumulative relation between two pluralities, i.e. to understand how the properties of the individual parts of the two pluralities add up to the properties of the whole in relation to each other, but also, and even in the absence of explicit marking, to form representations that involve a distributive universal quantifier. Likewise, the capacity to derive cumulative and distributive readings does not seem to rely on lexical knowledge, which may be symptomatic of its generality. Hence, altogether the findings of this study can contribute to our understanding of children's non-adult like responses in tasks that include plural and quantified sentences and help to reveal the mechanisms that guide the acquisition of logical expressions.

Let us come back to the question of why the priming effect we found was asymmetric and spell out in more detail how structural complexity might have affected our results. We saw that the priming effect of distributive readings was smaller than the effect of cumulative readings in that the difference between distributive conditions and baseline was smaller than the difference between cumulative conditions and baseline. Assuming that the cumulative interpretation is simpler and the distributive interpretation involves additional structure, as proposed by Link (1987), seems to be able to account for the pattern in the following way: During the priming phase in the distributive condition, a complex structure that involves the cumulative representation and the additional distributivity operator was activated, as schematized in (35). This facilitated re-activation of the distributive interpretation (35a), and at the same time of the simpler, cumulative interpretation (35b).

(35)   **Distributive prime:** [D [CUM]]

    a.   [D [CUM]]

    b.   [CUM]

During the priming phase in the cumulative condition, only the simpler cumulative structure, schematized in (36), was activated. The cumulative representation of the sentence accessed in the priming phase lead to an increased availability of cumulative interpretations in subsequent target trials (36b), however, its effect on the availability of the distributive interpretation was only limited since not all formal features that characterize the distributive interpretation (36a) were activated during priming.

(36)   **Cumulative prime:** [**CUM**]

    a.   [**D** [**CUM**]]

    b.   [**CUM**]

Therefore, the distributive structure might have primed both cumulative and distributive interpretations, leading overall to a smaller effect on distributive interpretations. The cumulative structure on the other hand primed cumulative interpretations, but might not have had a similar effect on the availability of distributive interpretations. This might partly explain why we found a smaller effect for the distributive interpretation, presupposing that multiple structural elements may be the source of priming effects.

One caveat concerning the scope of our investigation is that our results are not informative with respect to the question of whether preschoolers' capacities are restricted to linguistic representations of quantification and plurality. More precisely, it is not clear whether children deploy the corresponding concepts in thinking about multiple individuals, e.g. when encoding information or reasoning about multitudes, or whether this representational resources are only deployed given verbal input.

As some have pointed out, the details of the relationship between the semantics of natural language and logical thinking are underexplored (cf. Canudas-Grabolosa et al., 2023). Even if we assume that a compositional, symbolic thought-device, a LOT, exists, it is still possible that the meanings of natural language expressions are just reflexes of the corresponding expressions in LOT. The relationship may thus be characterized as a mere resemblance rather than an equivalence.[17] Let us formulate possible scenarios on how natural language and LOT might be

---

[17]This issue is related to but different from the question of whether linguistic and/or non-linguistic logical

embedded in cognition and how linguistic and non-linguistic logical representations might be related:

(I) One possibility is that all symbolic, logically structured representations are generated by the language system, rendering complex thought represented in a LOT derivative of and dependent on linguistic representations. On this view, the function of language is to create or enable thought, and we would expect that the expressive power of LOT equals the expressive power of natural language.

(II) Another possibility is that the system that generates complex thought exists prior to language and linguistic representations are derived from LOT representations. According to this view, the function of language is to express thoughts, and we would expect that the expressive power of LOT is equal to or greater than that of natural language.

(III) Yet, another way to think about this relationship would be in terms of two independently evolved systems.

Depending on one's view of this relationship, this might be compatible with one of the following hypotheses about linguistic and non-linguistic logical primitives:

(i) Natural language and LOT share a common toolkit of logical primitives. This hypothesis seems plausible given the assumptions formulated in (I) and (II). Without further assumptions, it seems not consistent with the view described in (III).

(ii) Natural language and LOT differ in the logical primitives that may be deployed. If natural language and LOT represent two independent systems, as postulated in (III), it would not be unexpected that each system uses its own logical vocabulary. However, this might also be the case under the assumption in (II).

The questions surrounding this topic are ultimately empirical questions which can be approached experimentally, for instance by studying whether non-verbal and verbal reasoning

---

representations correspond to concepts from formal logic (for discussion see e.g. Crain, 2012; Canudas-Grabolosa et al., 2023).

involves the same logical structure. The next chapter presents findings from a study investigating whether reflexes of universal quantification can be observed in a non-verbal task with preverbal infants, thus offering an attempt to address some of the issues raised above.

Chapter 3

Quantification in infants

3.1 Introduction

In the previous chapter we have demonstrated that children have access to a distributive quantifier meaning before they have been shown to master the truth-conditions of overt distributivity markers and we have argued that these findings support the Mapping Hypothesis. As a next step, we set out to explore whether quantification is an operation only present at a linguistic level or whether it can be found that linguistic and non-linguistic representations share core formal properties that characterize quantification. The study described below investigated whether already preverbal infants can quantify over small sets of individuals when reasoning about the outcome of an event that involves multiple agents.

Despite there being initial evidence that infants possess an early capacity to deploy logical concepts, for instance learning a rule based on the concept SAME (but not DIFFERENT) (e.g. Hochmann et al., 2016, 2018; Hochmann, 2022) and drawing disjunctive inferences (Bohus et al., 2023; Cesana-Arlotti et al., 2018, 2020; Ekramnia et al., 2021), little is known about infants' capacity to form quantified representations. Besides a few conference presentations there are no published studies that dealt with this topic.

Téglás and Bonatti (2009) presented one such attempt, reporting findings from a violation-of-expectation study with 10-month-olds which concluded that infants may encode the logical structure of simple scenes. In these experiments, infants initially saw a scene with two doors and an unfamiliar agent that repeatedly opened each door. These scenes gave infants two options. They could encode the events by giving priority to the numerical information (i.e. two doors were opened) or encode them in terms of their quantificational properties (i.e. all doors were opened). At test, infants saw two types of events which this time involved three doors, in one the agent opened all, in the other one only two of them. It was this last event structure that elicited longer looking times, indicating that infants were expecting the agent to open all of the doors, irrespective of the number of doors.

The findings from a recent study with infants of the same age by Cesana-Arlotti, Knowlton, Lidz, and Halberda (2020) suggest that the ability to distinguish between actions that are performed individually and collectively by an exhaustive set of agents is available from early on. In this habituation experiment infants were presented either with collective or individual chasing events in form of animated movies displaying dynamic scenes that involved three chevrons and three three balls. In collective chasing events the chevrons were chasing one ball together, while in individual chasing events each of them was chasing one ball individually. In the test phrase, infants saw either the same type of event or the other one. Those infants who were habituated to collective events dishabituated to individual events and those who were habituated to individual events dishabituated to collective ones, suggesting that they distinguished between exhaustive collective and individual actions, thus possibly might be attuned to the difference between quantifiers such as 'each' and 'all'.

As outlined earlier, it is a debated question how children acquire concepts such as ALL.[1] Following advocates of the Learning Hypothesis, there is a discontinuity between infants' and adults' conceptual system in that the former is a representational system that lacks the notions of set and quantification. The assumption is that it is through language that such concepts

---

[1]Here and throughout the text we use small caps to indicate that we are talking about concepts which are similar but not necessarily equivalent to the underlying meanings of linguistic expressions such as *all*, *some* and *not*, and do not necessarily correspond to the logical operators $\forall$, $\exists$ and $\neg$.

become available, more precisely, that the acquisition of plural morphology and quantifying expressions enriches the initially less powerful system by making available new representational resources. Thus, language is argued to enable children to form symbolic set-level representations (cf. Carey, 2009, 2011). The Mapping Hypothesis on the other hand assumes a developmental continuity in that infants' system has already all the representational capacities that characterize the system of adults.[2] Language does not alter the representational resources qualitatively, but rather exploits them. Therefore the initial conceptual system will already have the capacity to form symbolic representations of sets and quantifiers (cf. Fodor, 1975). The crucial difference between the Learning Hypothesis and the Mapping Hypothesis is then that the former claims that the representations which become available through language are qualitatively different from, and not even expressible in terms of, the previously available representations, whereas the latter states that language builds on antecedently existing representations, possibly by making certain distinctions salient and/or referencing to parts or combinations thereof. Our study was designed with the aim to test these claims by looking at infants' early capacity to deploy universal quantification. We focused on the logical concept of *exhaustivity*, a concept which is part of the meaning of universally quantifying expressions and requires set-level representations.

3.2 Exhaustivity in language development

Research on language development has shown that, in addition to quantifier spreading and difficulties with restricting the meaning of distributive quantifiers, preschool-aged children frequently exhibit what is known as the *underexhaustive error*. Consider for instance the sentence in (37).

(37) *Every girl is holding a puppet.*

CEU eTD Collection

---

[2]In Carey (2009) these hypotheses are called accordingly the *Discontinuity Thesis* and the *Continuity Thesis*.

(38)   There are three girls, Sue, Ann and Marie.

   a.   EXHAUSTIVE SCENARIO: Ann is holding one puppet, Sue is holding one puppet, and Marie is holding one puppet.

   b.   NON-EXHAUSTIVE SCENARIO: Ann is holding one puppet, Sue is holding one puppet, Marie is not holding anything.

For adults, such a statement is true in the exhaustive scenario in (38a), in which all of the relevant girls are holding a puppet, and false in the non-exhaustive scenario in (38b), in which there is a girl who is not holding a puppet. Accordingly, a standard assumption is that (part of) the semantic contribution of universal quantifiers such as *every*, *each* and *all*, irrespective of the distributive/non-distributive component, is that they carry universal force, requiring that there are no exceptions in the set of girls with respect to the property of holding a puppet (e.g. Barwise and Cooper, 1981). However, several studies have documented that children sometimes accept sentences like (37) even in non-exhaustive scenarios such as (38b) (cf. Inhelder and Piaget, 1964; Freeman and Stedmon, 1986; Philip, 1995; Roeper and Pearson, 2004; Altreuter and de Villiers, 2006; Aravind et al., 2017; Drozd et al., 2019, a.o.; see also section 2.3).

For instance, such findings are reported by Altreuter and de Villiers (2006), who conducted a study with English-speaking 5-to-7-year-old children examining both comprehension and production of the universal quantifier *every*. In the comprehension task, children were presented with sentences like *Every cat has a cookie* and asked whether the sentence matched an accompanying picture. The test pictures depicted a range of scenarios, including exhaustive scenes which involved extra elements (e.g. three cats each holding a cookie with an extra cookie present, or three cats each holding a cookie alongside a sheep holding a banana) and non-exhaustive scenes (e.g. three cats each holding a cookie and one cat holding a banana). On the following day, the same children participated in a production task in which they were shown similar pictures and asked to produce a sentence beginning with *every* that accurately described the scene. These pictures were slightly modified to allow participants to formulate a truthful universal statement (e.g. in non-exhaustive scenarios where only three cats were holding a cookie and one was holding a banana, all four cats were wearing hats). Relevant here is

57

that, unlike errors observed with other types of scenarios, underexhaustive errors in comprehension were found primarily in the youngest group, i.e. 5-year-olds sometimes accepted such sentences in non-exhaustive scenarios. Notably, these children also produced universally quantified sentences given non-exhaustive scenarios, suggesting that the underexhaustive pattern extended beyond comprehension into production.

In a longitudinal study, Aravind et al. (2017) replicated these findings on children's comprehension on sentences with *every* and further demonstrated a developmental dissociation between underexhaustive errors and quantifier spreading. Specifically, they found that underexhaustive errors are primarily observed in younger children, while spreading errors emerge later and may even persist into adulthood. Participants were English-speaking children tested four times between the ages of 4 and 7, using sentences such as those in (39a) and (40a) alongside pictures depicting either non-exhaustive scenarios (39b) or scenarios with an extra object (40b).

(39)　a.　*Is every woman sailing?*

　　　b.　NON-EXHAUSTIVE SCENARIO: There are four women. Three of them are each sailing a boat, one woman is sitting on the beach.

(40)　a.　*Is every daddy holding a baby?*

　　　b.　SPREADING SCENARIO: There are three daddies. Each is holding a baby, an additional baby is sitting in a buggy.

At around age 4, most children (100 out of 140) provided correct responses to universal sentences in spreading scenarios, but only 22 did so in non-exhaustive scenarios. Accuracy in non-exhaustive scenarios improved with age, 63 children answered correctly at age 5, 85 at age 6, and 126 at age 7. In contrast, performance in spreading scenarios declined, only 64 children at age 5 responded correctly on all trials, 43 at age 6, and 89 at age 7. Spreading errors were also more frequent than underexhaustive errors in an adult control group with participants providing correct responses in 76% of spreading trials versus 93% in non-exhaustive trials.

Interestingly, a large-scale cross-linguistic study by Drozd et al. (2019) which systematically investigated children's interpretation of universal statements across 12 languages (Catalan,

Cypriot Greek, Danish, Dutch, German, Croatian, Polish, Russian, Serbian, Slovak, Lithuanian and Maltese) found a uniform pattern of errors. 4-to-5-year-old children were presented with sentences such as *All the horses are eating an apple* accompanied by non-exhaustive scenarios which featured an additional character not performing the respective action, scenarios with an extra object, and scenarios containing an extra character performing a different action on a different object, as in the study by Altreuter and de Villiers (2006). With few exceptions, children across languages showed similar error rates across these scenario types. The highest error rate occurred in scenarios with extra objects (58%), followed by scenarios with extra characters and objects (44%), and the lowest in non-exhaustive scenarios (11%).

Although these rates do not perfectly match findings from previous studies (e.g. Aravind et al., 2017), the results nevertheless show that difficulties with universally quantified sentences are not language-specific. Rather, children acquiring a range of typologically diverse languages appear to struggle with the same types of challenges. Overall the findings support the view that quantifier spreading errors may stem from extra-linguistic factors, with even adults sometimes being misled by contexts that involve extra objects (cf. also Crain et al., 1996; Brooks and Sekerina, 2006; Drozd and Loosbroek, 2006; Philip, 2011, a.o.). The developmental trajectory of underexhaustive errors together with the parallel between children's performance in comprehension and production of *every* on the other hand suggests that younger children may not yet fully grasp the link between universal quantification and exhaustivity, and only gradually learn to interpret universals exhaustively. Altreuter and de Villiers (2006) propose that younger children may interpret *every* as meaning something like *a lot of* and, along similar lines, Aravind et al. (2017) suggest that there may be a developmental stage at which children assign a plural existential quantifier meaning to the universal. A key puzzle about children's language development then is how they come to know that universals like *each*, *every* and *all* enforce exhaustivity.

59

3.3 The representation of exhaustivity in preverbal infants

We considered the possibility that exhaustivity might be a primitive concept deployed in thought even by preverbal infants. We hypothesized that such a concept might be entertained for delimiting the hypothesis space when infants learn about their environment by hypothesis testing. It seems that for learning to be successful it would be crucial to formulate a set of hypotheses and to exhaustively perform hypothesis testing on this set. Similarly, without considering an exhaustive set of available alternatives infants probability intuitions (cf. Téglás et al., 2011) would generate unreliable expectations.

In our study we therefore aimed to find out whether infants can distinguish between exhaustive and non-exhaustive scenes and whether they can derive a rule-based prediction based on this contrast. Exhaustivity was operationalized as an action that was performed by multiple agents whereby either all agents participated in the action or only a subset did, corresponding to descriptions that involve the concepts ALL and NOT ALL, respectively. If the Mapping Hypothesis is correct, infants should have access to concepts such as ALL even before the onset of language and should be able to apply those to sets of varying cardinality. In contrast, according to the Learning Hypothesis, preverbal infants only can track and individuate up to three (sometimes four) objects but lack the ability to represent set-level features which are independent of individual objects. Thus, such a system would allow infants to encode the number of agents involved in an action (e.g. two or three) but would not allow them to represent that ALL agents, regardless of their number, participated.

If 12-month-old infants can learn a rule based on the quantificational properties of the scenes, this would suggest that they, before even having acquired the singular/plural distinction, are able to form abstract representations of the presented scenes which involve universal quantification. Success at that age in a non-verbal task would further hint at common formal properties shared by linguistic and non-linguistic representations. If, however, infants lack the representational capacities to form such representations, we would not expect them to be able to learn the rule. Such a failure might be explained by such representations being language-dependent in

the sense that they are only learned once the corresponding linguistic expressions are acquired. Of course even in this case, quantification might be a mental operation that in this form is only available at a linguistic level of representation.

We developed an anticipatory eye movement paradigm in which infants watched video animations showing multiple agents acting upon an object in a goal-directed manner. The scene mimics a collaborative interaction, where the joint effort of the protagonists leads to a successful displacement of an object. The collaborative actions we selected for our study gave ample possibilities to manipulate the quantificational structure necessary to grasp the events and provide an appropriate description of the scene.

Numerous studies have demonstrated that infants can track and recognize various instances of goal directed behavior from very early on (Gergely et al., 1995; Gergely and Csibra, 2003; Kamewari et al., 2005; Csibra, 2008, a.o.). Their ability to establish the goal of an action, even if the behaviors they observe are unfamiliar to them or are not yet included in their own motor repertoire, suggests the presence of a powerful inferential device at an early age. Goals, however, can be attributed both to individuals and to collections of individuals. At around 6 months of age, infants start tracking the size of the social groups (e.g. when they have conflicting goals, Pun et al. 2016), by 9 months of age, they represent simple forms of collaboration (Begus et al., 2020), and by 18 months of age, they have a solid understanding of how actions performed by different agents complement each other to contribute to a hierarchical goal structure (Fawcett and Gredebäck, 2013, 2015; Henderson and Woodward, 2011; Henderson et al., 2013). Infants do not only enter into cooperative activities (Warneken and Tomasello, 2007), but they also expect that others will do the same (Vizmathy et al., 2024). These findings suggests that infants can process and encode various dimensions of collaborative interactions in their first year of life starting from the number of agents involved to their shared goals.

In the current study, the two conditions of the learning phase implemented the quantificational structure by manipulating the relation between the number of agents present in the scene and the number of agents actually contributing to the goal. In the exhaustive condition, all agents (3 out of 3) collaborated towards the goal, whereas in the non-exhaustive condition, their con-

tribution was not exhaustive and only some of the agents (2 out of 3) teamed up. While in typical collaborative tasks success often hinges on the number of partners involved, in this case, achieving the goal was not influenced by their number. Crucially, each condition was predictive of a different outcome. Upon achieving the goal, one of the two laterally positioned objects started rotating, for instance the left one moved after an exhaustive and the right one moved after a non-exhaustive scene. This is the predictive relation that we expected infants to learn. Anticipatory eye movements served as an implicit measures of the targeted cognitive processing. We selected these outcomes to remain neutral regarding the social structure of these scenes (i.e. there is nothing in these events that would signal a causal relation with the exhaustive or non-exhaustive nature of the observed collaborative acts).

This procedure, originally developed to study infants' categorization (McMurray and Aslin, 2004), was successfully adapted to various other domains, like speech perception (Albareda-Castellot et al., 2011), rule learning (Kovács and Mehler, 2009; Hochmann et al., 2011, 2018), relational concepts (e.g. SAME and DIFFERENT, Addyman and Mareschal 2010; Kovács 2014; Hochmann et al. 2018) and the role of communicative cues (Kovács et al., 2017). Only the efficient processing of the cue characteristics and their reliable discrimination in the learning phase enables infants to develop predictions congruent with the reward contingencies. The key signature of successful discrimination is the performance the participants produce at the and of the learning phase. The patterns produced in the test phase are expected to reveal the stimulus properties that infants prioritized in their generalization. Infants failure to discriminate eventually will be reflected by the lack of anticipatory eye movements at all stages of the experiment.

In the learning phase of our experiment, infants could learn to predict outcomes by relying on either the number of agents involved in the pushing action (2 vs. 3) or the quantificational properties of the scenes (non-exhaustive vs. exhaustive). During the test phase, infants observed events which involved 2 out of 2 agents. These events resembled the exhaustive learning trials in that all available agents participated in the action, but they matched the non-exhaustive trials in terms of absolute set size. We hypothesized that if infants had encoded the number of agents

during the learning phase, they would look toward the object associated with non-exhaustive events (i.e. when 2 out of 3 agents acted). By contrast, if infants had encoded the quantificational structure of the events, they should look toward the object associated with exhaustive events (i.e. when all agents in the set, meaning 3 out of 3, participated in the action).

3.4 Experiment 3

3.4.1 Methods

This experiment was preregistered (https://aspredicted.org/fjpm-76mh.pdf).

Participants

Our final sample included 47 typically developing, full-term 12-month-old infants who, depending on their learning performance (see the description of the learning criteria in the results section), were split into two subgroups. There were 24 learners (M = 12;01; SD = 7.5 days, 11 females) and 23 non-learners (M = 12;03; SD = 7.3 days, 10 females). Another 40 infants were tested but had to be excluded due to fussiness or inattention (n = 23), crying (n = 6), technical failure (n = 3), parental intervention (n = 1) or for not providing sufficient gaze data (n = 7; see criteria for data quality in section 3.4.2).

We initially established the sample size following prior studies investigating rule-learning in 12-month-old infants with a similar paradigm (Hochmann et al., 2011, 2018) supported by a power analysis indicating that the sample size adopted in these studies (N = 24) yielded adequate power (> 0.8). Based on pilot data suggesting that the quantifier rule was not dominant in that age-group, we decided to split the sample into learners and non-learners and stopped data collection when we reached 24 learners.

The testing sessions took place at the CEU Babylab (Budapest, Hungary) and all recruited

infants received a small toy as a gift for participation. The study was approved by the United Ethical Review Committee for Research in Psychology in Hungary (EPKEB) and informed consent was obtained from the participants' caregivers before the testing session.

Stimuli

The stimuli were 3D animations created with Autodesk Maya animation software version 2020.4 (Autodesk, 2020). The resulting movies were played at 25fps. We prepared three kind of movies: Two served as stimuli for the learning phase and one was displayed during the test phase.

Each trial of the learning phase started with a static scene presenting a centrally positioned rectangular area defined by a low wall (1800 ms). The front and back walls of this structure were parallel to the lower and upper edge of the screen and had an opening (15,5 cm) in the front (45 cm). This area was flanked on its left and right corner by two multicolored, cube-shaped objects (occupying a 4,8 x 4,4 cm area on the screen). Then three agents entered the scene one by one accompanied by sounds (4200 ms). While only minimally defined, the pink spherical objects (having a diameter of 3.8 cm) that served as agents were self propelled and demonstrated coordinated, goal directed actions. In exhaustive learning trials, all of the three present agents jointly pushed a green box towards the central enclosure and stopped after successfully carrying their load across the entrance (5000 ms). After the completion of this action, there was a short pause (1500 ms) followed by a rotation of one of the target objects (2000 ms) and a longer pause after the rotation had stopped (2000 ms). There was a sound marking the onset of the target motion. We assumed that these visual events will be highly engaging for infants, motivating them to orient their gaze to their location eventually developing predictive eye movements in their direction before their moving starts. In non-exhaustive learning trials, only two out of the three agents jointly pushed the box towards the central area, followed by a rotation of the other target object.

Test videos were similar to those used in the learning phase, but differed in two respects.

Whereas the events in the learning phase involved three agents, test trials featured only two agents. After entering the scene (5200 ms), the two agents pushed together a green box to the center of the enclosure (5000 ms). Unlike in the learning trials, this was followed by a static scene (4000 ms), in which the targets did not move.

Procedure

The experiment had two phases. In the learning phase participants were presented with 12 trials. The subsequent test phase comprised 6 test trials.

Learning phase In the learning phase, trials consisted of movies that involved three agents. In all trials, agents were pushing a box together towards a goal, after which one out of two laterally placed target objects started to rotate. In half of those trials, the set of agents involved in the pushing action was exhaustive whereby all three agents contributed to the joint goal achievement. In the other half of the trials, the set of pushing agents was non-exhaustive whereby only two of the three agents participated in the joint action. The exhaustive/non-exhaustive character of the event was predictive of the rotation of the target objects (Figure 3.1). The side of the correct target object, which in our study was the object that started to rotate in exhaustive events, was counterbalanced across participants. The presentation of the exhaustive/non-exhaustive trials was performed in three blocks and followed either an ABAB-BABA-ABBA or an AABA-BBAB-AABB order. The sequences allowed for one repetition of the same condition in consecutive trials. The last block of all sequences was such that it contained an equal number of trials per event type.

Test phase In the test phase, the presented events were similar to the exhaustive events seen during the learning phase, except that they exhibited different numerical properties. This time only two agents instead of three were present and both contributed to the joint goal. Thus, the events shown in test trials were similar to exhaustive learning trials in that all agents participated in the action, but in terms of the cardinality of the set of agents, they were similar to non-exhaustive learning trials. During the learning phase it was possible to consider either the

65
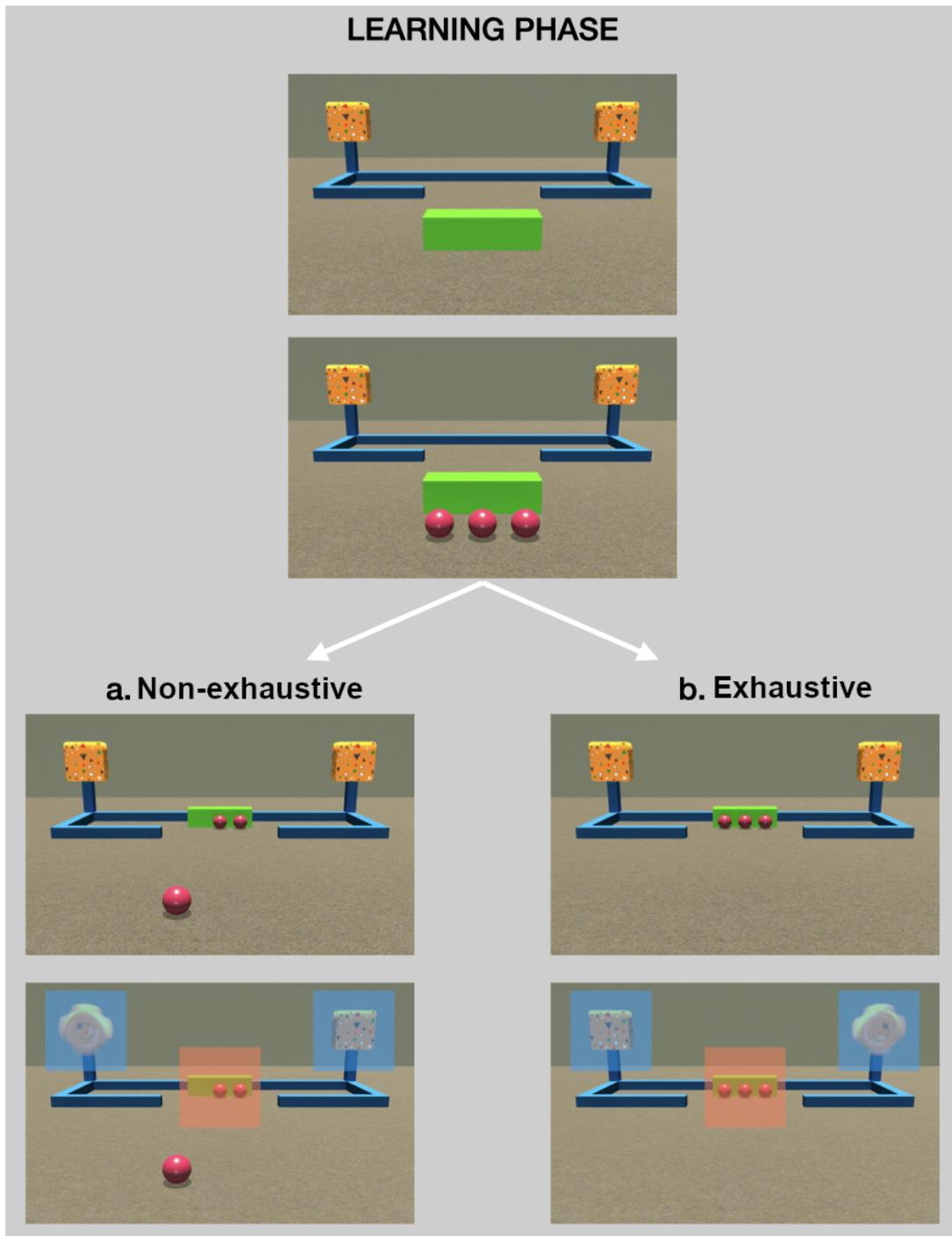
Figure 3.1: Event sequences presented in the two conditions in the learning phase. After the agents enter the scene and their cardinalities can be encoded infants are presented with a) *non-exhaustive scenes*, where 2 out of 3 agents contribute to a joint goal, and b) *exhaustive scenes*, where 3 out of 3 agents' collaborative actions can be seen. a-1) and b-1) The spatial configurations marking the onset of the time-window for anticipation. a-2) and b-2) The rotation of the target and the AOIs. The central AOI is marked with a red rectangle, the lateral target AOIs with blue rectangles.

number of the acting agents (2 vs. 3) or the set-level properties (NOT ALL vs. ALL) to predict

the side of the rotating target object. Changing the number of agents involved in the test scenes

66

aimed at detecting the rule that was generalized by infants. If infants generalized a rule based on the exhaustivity of the set, we expected them to look toward the target object that rotated in the learning phase after all agents pushed the box. If, however, infants formed expectations based on the numerical properties of the scenes, we expected them to look at test toward the object which rotated after only two out of three agents pushed the box in the learning phase. In test trials the agents' goal achievement did not result in a rotation of the lateral objects. The time-window corresponding to this static scene at the end was used to collect our dependent variables and to verify the generalization. The dependent measure was infants' predictive eye movement (first looks and cumulative looking time) towards the areas of interest (AOIs) containing the target objects.

Participants were tested in a darkened and soundproof room. They sat on the parent's lap, approximately 60 cm away from the monitor. Parents wore opaque sunglasses and were instructed not to interact with the infant during testing. Infants' gaze was recorded at 60 Hz with a Tobii T60XL eye-tracker (http:// www.tobii.com). The stimuli were presented on a 24-inch screen integrated with the eye tracker (resolution: 1920 × 1200). Stimulus presentation and gaze data collection was controlled by the software PsyScope X (http://psy.cns.sissa.it/) running on an Apple Mac Mini 2,8 GHz Intel Core i5. The experiment was preceded by a five-point calibration procedure. Before each experimental movie, a salient visual attractor was presented on the screen to help infants to reorient their attention toward the stimuli.

3.4.2 Results

Inclusion criteria All infants submitted to the analysis contributed with at least 4 valid points to the eye tracker's calibration. The analysis was run over gaze coordinates averaging the data collected from both eyes when it was available, otherwise data from one eye was used. In order to guarantee the reliability of the data, we applied both trial-level and infant-level inclusion criteria, which were established prior to the experiment and pre-registered. A trial was excluded from the analysis if infants provided less than 50% gaze data in the relevant time-
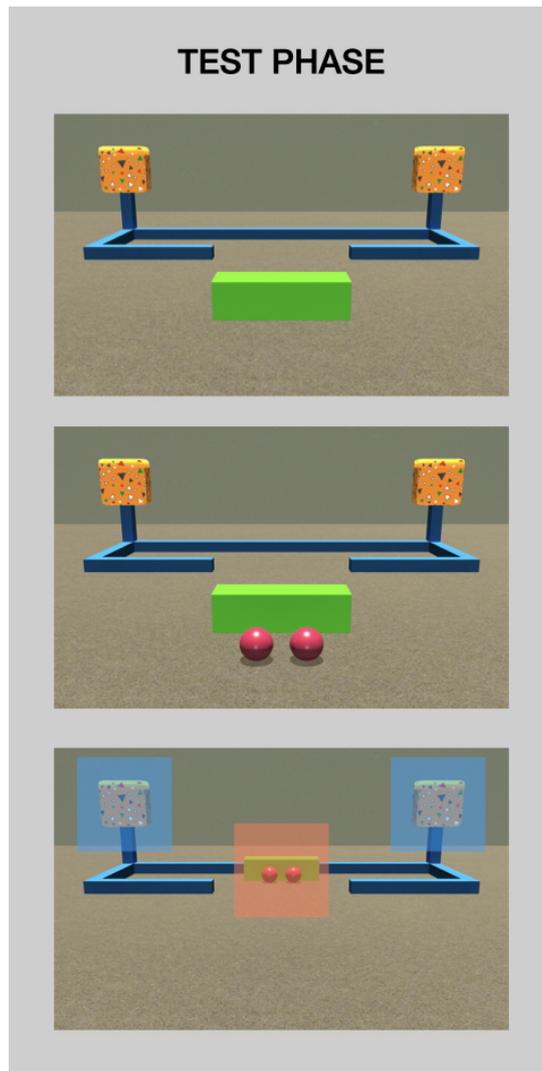
Figure 3.2: Event sequences presented in the test phase. After two agents enter the stage and achieve their joint goal the scene remains static for 4000 ms. This time-window serves to collect our dependent variables, infant's first looks and cumulative looking time towards the areas of interest (AOIs) containing the target objects. The central AOI is marked with a red rectangle, the lateral target AOIs with blue rectangles.

window (i.e. the last phase of the test trial) and, additionally, no gaze shift from the central AOI towards the target AOIs. Each AOI was 400×400 pixel large. The target AOIs comprised the region around the target objects. The central AOI captured the static agents at their final destination. Infants were excluded from the analysis if they had fewer than 3 test trials[3] that met the inclusion criteria ($n = 4$). All analyses were performed in R (v4.0.5, R Core Team, 2021).

---

[3]In the preregistration this was set to 2 test trials, however, since there were 6 test trials in total, we decided to apply a more conservative measure for ensuring data quality and representativeness.

Learning criteria Based on their performance in the final block of the learning phase (which encompassed two exhaustive and two non-exhaustive trials), participants were classified into two subgroups. Their performance was analyzed in the form of standard difference scores. For each trial the time infants attended to the incorrect, i.e. rule incongruent, target AOI (IC) was subtracted from the gaze duration collected from the correct, i.e. rule congruent, target AOI (C) and the result was divided by the total amount of looking collected from both target AOIs (df = (C-IC)/(C+IC)). The resulting measure ranges from -1 to 1, the positive values expressing rule-congruent predictions. Infants who looked more to the correct than to the incorrect side were classified as learners (overall: $M = 0.56$, $SD = 0.36$; in exhaustive trials: $M = 0.60$, $SD = 0.58$; in non-exhaustive trials: $M = 0.47$, $SD = 0.79$), infants who looked more to the incorrect than to the correct side were classified as non-learners (overall: $M = -0.65$, $SD = 0.31$; in exhaustive trials: $M = -0.70$, $SD = 0.51$; in non-exhaustive trials: $M = -0.58$, $SD = 0.58$).[4] Participants who did not look to any of the target AOIs in the final block of the learning phase were excluded from the subsequent analysis ($n = 3$).

Data analysis The principal dependent variables, designed to capture whether infants could generalize the rule to new cardinalities (i.e. 2 out of 2 agents involved in the action) that were not present in the learning phase (where 3 out of 3 agents were involved in the action), were derived from infants' predictive eye-movements towards the target AOIs. To find out whether infants predict the correct location of the target based on the quantifier learned in the preceding phase, we measured i) first looks and ii) cumulative looking time towards the AOIs including the target objects. Difference scores were calculated for each measures and tested against chance. The first looks were those eye-movements that started from the central AOI and shifted towards a target AOI after the agents had stopped. To get the difference scores over the first looks for each infant, we subtracted the number of trials where the first saccade occurred to

---

[4]The performance of learners and non-learners was not only consistent across trial types but also quite consistent across blocks: Learners looked more to the correct than to the incorrect side in the first block (overall: $M = 0.31$, $SD = 0.75$; in exhaustive trials: $M = 0.30$, $SD = 0.86$; in non-exhaustive trials: $M = 0.32$, $SD = 0.96$) and in the second block of the familiarization (overall: $M = 0.42$, $SD = 0.73$; in exhaustive trials: $M = 0.23$, $SD = 0.75$; in non-exhaustive trials: $M = 0.60$, $SD = 0.87$). Non-learners showed a slight preference for the correct side in the first block (overall: $M = 0.20$, $SD = 0.83$; in exhaustive trials: $M = -0.02$, $SD = 0.82$; in non-exhaustive trials: $M = 0.70$, $SD = 0.95$), but started to look more to the incorrect side from the second block onwards (overall: $M = -0.31$, $SD = 0.89$; in exhaustive trials: $M = -0.26$, $SD = 0.94$; in non-exhaustive trials: $M = -0.38$, $SD = 0.87$).

the incongruent AOI from the number of trials with first saccade to the congruent AOI, then we divided the result by the total number of trials where there was a saccade to the target AOIs. Difference scores based on cumulative looks were calculated in the same manner as those described in the paragraph outlining the learning criteria.
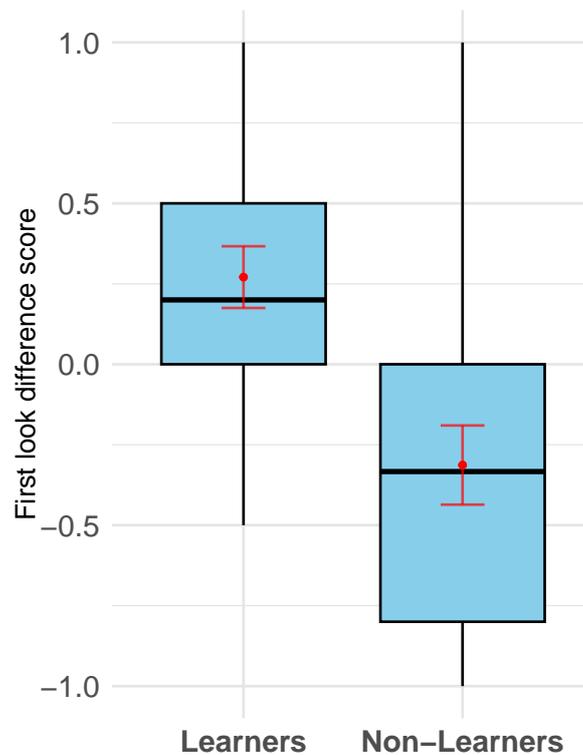
Figure 3.3: **First looks** Difference scores of first looks for learners and non-learners. Horizontal lines mark the medians and boxes represent the upper and the lower quartiles. The whiskers extend to the smallest and largest data points within the $1.5 \times$ interquartile range from the upper and lower edges of the middle quartiles. The red diamonds show the mean values of each group and the error bars represent the standard errors.

First looks Figure 3.3 shows the first looks of learners and non-learners towards the congruent target AOIs operationalized as difference scores. Learners produced more first gaze shifts towards the congruent target AOI than towards the incongruent target AOI ($M = 0.27$, $SD = 0.47$), whereas for non-learners the opposite was true ($M = -0.31$, $SD = 0.59$). Infants' performance in the learners group was significantly above chance, $t(23) = 2.82$, $p = .01$. In contrast, infants in the non-learners group produced less looks to the correct side than expected by chance, $t(22) = -2.55$, $p = .018$.

Cumulative looking time As illustrated in Figure 3.4, learners showed longer looking times
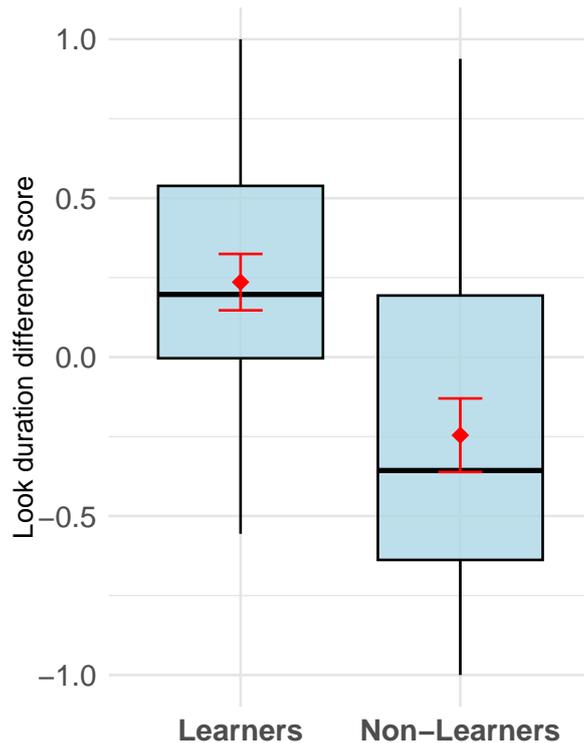
Figure 3.4: **Cumulative looking time** Look duration difference scores for learners and non-learners with the horizontal lines representing the medians and boxes indicating the upper and the lower quartiles. The whiskers show the smallest and largest point within the $1.5 \times$ interquartile range from the upper and lower edges of the middle quartiles. The red diamonds mark the mean value of each group with the error bars indicating the standard error.

to the congruent target AOI than to the incongruent target AOI ($M = 0.24$, $SD = 0.43$), which was significantly above chance, $t(23) = 2.66$, $p = .014$. Non-learners on the other hand looked longer to the incongruent target AOI than to the congruent one ($M = -0.25$, $SD = 0.55$) and spent less time at the congruent target AOI than expected by chance, $t(22) = -2.12$, $p = .045$.

3.4.3 Discussion

Our findings are consistent with the idea that 12-month-old infants can distinguish between exhaustive and non-exhaustive scenes and possess the capacity to learn a rule based on quantificational properties of an event, though this may not be their dominant way of encoding events that involve multiple agents. Based on their performance in the learning phase of the experiment we identified two groups of participants who exhibited a different looking behavior in our task.

The predictive eye-movements of infants in the learners group in test trials suggest that they expected that an action of an exhaustive set of agents would lead to an outcome that was linked to an action of an exhaustive set of agents during familiarization, irrespective of the difference in numerical properties. Learners not only produced significantly more first looks towards the side that was congruent with a rule that involved ALL than expected by chance, they also looked significantly longer to the correct side, i.e. to the object that was supposed to rotate given an exhaustive set. Their performance in the test phase closely matches their performance in the learning phase, where they anticipated outcomes of both exhaustive and non-exhaustive events with a comparable accuracy. The results of the non-learners on the other hand show that they did not form similar expectations and did not base their anticipatory looking behavior on such a rule. There were significantly more first looks towards the rule-incongruent side in the non-learners group when presented with exhaustive events at test. Similarly, non-learners looked significantly longer to the incorrect side, i.e. to the object that was linked to non-exhaustive events during familiarization. However, this does not necessarily mean that non-learners formed expectations based on the numerical properties of the scenes, since this trend was already observable during familiarization.

The overall pattern is compatible with two conclusions about the underlying reasoning processes of learners and non-learners: For one, learners may have taken into account the contrast between exhaustive and non-exhaustive sets of agents (ALL agents vs. NOT ALL agents). Second, this set-level information might have served as a premise to form a rule with a particular outcome. If they had focused on the numerical properties of the familiarization events instead, in which the exhaustive events involved an action performed by three out of three agents and the non-exhaustive events involved an action performed by two out of three agents, they would have predicted the opposite outcome at test, where two out of two agents were pushing the box. Interestingly, infants in the non-learners group might have also considered the exhaustive/non-exhaustive contrast. Their looking behavior at test aligned with their looking behavior in exhaustive learning trials, suggesting that events with an exhaustive set of agents were treated alike. However, we did not find evidence for them connecting this information with the outcome of the event and deriving a prediction about the correct location of the ro-

tating object. Possibly infants in the non-learners group paid more attention to other aspects of the scene, e.g. the collective aspect of the action or the goal-directedness, and thus did not form rule-based expectations. In this sense our results can be interpreted as indicating that at 12 months of age infants are in a developmental stage where the categorical distinction between exhaustivity and non-exhaustivity is not considered relevant – at least when encoding events of the type that we presented to them. Changing certain parameters or introducing other types of actions could eventually affect infants' encoding strategies and render an understanding in terms of quantification more natural, e.g. non-social contexts or contexts that involve a number of agents that is beyond subitizing range, in which case lacking access to the exact cardinalities may prioritize descriptions in quantificational terms.

Only about half of our valid participants picked up the appropriate contingencies during the learning phase of the experiment and correctly anticipated the side of the reward based on the stimuli that cued their prediction. While the paradigm used in our study had several advantages, certain features may have hindered infants' learning performance. One such factor was the fixed length of the learning phase. Each condition was represented by six trials played for the participants in a pseudo-random order. The preset trial number may underestimate the exposure slow learners may need to our stimuli and may overestimate the length of the learning phase for fast learners who might start to disengage. Either way, an implementation of this paradigm that can compute and track online whether a learning criterion was achieved during the learning phase may increase the sample of successful learners (cf. Shukla et al., 2011). Another factor that might have caused difficulties is that the paradigm requires that infants learn two rules, which may be too taxing for their learning system. In various applications of this paradigm, it was observed that infants have a tendency to generalize only one rule (the one that is simpler), for instance in studies that tested infants' ability to represent relational concepts like SAME and DIFFERENT (Kovács, 2014; Hochmann et al., 2011, 2018) 7- and 12-month-old infants learned only the simpler rule, in this case the one that involved the concept SAME.

Given the specifics of the implementation used in the learning phase, the quantification based approach is not the only one that can account for our results. The results can be accounted

for by at least two alternative explanations. The first alternative relates to the kind of events we chose. For instance, in non-exhaustive scenes, most agents participate in the collaborative action, while one agent does not. Such a bystander is absent from the exhaustive scenes. Infants could have exploited this difference by encoding the events in terms of the presence/absence of a bystander and base their predictions on this information. Though we cannot exclude this possibility, this sort of reasoning might be characterized by a degree of complexity that is comparable to the complexity of the underlying reasoning processes that we have described. Yet, another more parsimonious explanation is that infants connected the outcome of the events to low-level features of the scenes such as the distribution of agents. Indeed, at the end of the non-exhaustive scenes, the agents form a spatial configuration that is spread over a larger area of the stage (i.e.two agents are in the center while a third one is further away), while this is not the case for the exhaustive scenes (i.e. all agents are in the center). It must be left to future research to find decisive evidence for or against those explanations.

3.5   General discussion

From early on, infants have been shown to be sensitive to number. There is ample evidence that already very young infants are able to track and individuate up to three objects and to discriminate approximate quantities (a.o. Wynn, 1992a; Simon et al., 1995; Xu and Spelke, 2000; Feigenson et al., 2002; Feigenson and Carey, 2003, 2005; McCrink and Wynn, 2004). Results from a study by Wynn et al. (2002) suggest that 5-months-olds are also able to individuate and enumerate collective entities as they were shown to dishabituate to two vs. four collections of up to four items, thus also were able to group these into wholes. And in a manual search task, 14-month old infants were even able to hold four objects in parallel in working memory when chunking of objects into groups was made available through perceptual or conceptual cues (Feigenson and Halberda, 2008). These early capacities have been argued to rely on knowledge structures that are simpler in format and limited in scope, such as the object-tracking system and the analog magnitude system. Abstract representations of objects and sets

of objects are believed to emerge later, in tandem with increasing language competence (cf. Carey, 2009, 2011).

Our findings indicate that this picture might be incomplete, although they allow for multiple interpretations. On one possible interpretation, the results suggest that 12-month-olds are in principle able to abstract away from particular numerical information and encode events in terms of their quantificational structure. Infants in the learners group may have considered properties of sets rather than properties of individuals and may have based their predictions on such set-level information. Such a view does not necessarily challenge the existence of other, simpler knowledge systems that may be employed under certain circumstances, as indicated by the performance of the non-learners in our sample, who possibly encoded the scenes in terms of other features. Nevertheless, our findings could show that, in addition, infants possess more advanced representational resources, which enable them to reason and learn about sets of objects, and which ultimately may underlie their acquisition of logical expressions. Predictions in the learners group can be explained by infants having formed representations that involve the concepts ALL and NOT ALL, thus showing that infants at that age are sensitive to this distinction. The difficulties that children later encounter when interpreting universal quantifiers exhaustively would then not be due to them not being aware of this contrast, but rather not being able to map this distinction correctly onto linguistic forms.

If infants in the learners group formed representations based on the quantificational properties of the events as suggested above, it appears that these representations must have involved a contrast like ALL vs. NOT ALL since they learned to predict one outcome based on exhaustive sets, and, at the same time, a different outcome based on non-exhaustive sets. The relevant underlying distinction could not have been SOME vs. ALL, as describing the scenes with SOME would be compatible with both exhaustive and non-exhaustive scenes. Consequently, it would not account for infants deriving distinct predictions for both types of events. In fact it remains unclear whether infants' representation of non-exhaustive scenes involved an existential component, i.e. SOME $\wedge$ NOT ALL, since we did not present them with scenes in which none of the agents participated in the pushing action. Thus, we cannot determine whether their representa-

tion of non-exhaustive events would have been compatible with such scenes. Moreover, at this point it is important to note that, even if infants formed abstract representations of the logical structure of the different event types, this does not necessarily imply that those representations involved universal quantification. For instance, it is also conceivable that they simply encoded that exhaustive events featured only a pushing action, whereas non-exhaustive events included a pushing and a standing action.

Returning to the explanation based on the concept of a bystander, i.e. the idea that, rather than quantifying over sets, infants classified the scenes according to the presence or absence of an individual not participating in the collective action, it seems that such a representation could in principle involve the same logical concepts. In particular, the detection of a bystander could rely on a representation of a collective action in which not all agents participated, compared to a collective action in which all agents took part. However, a bystander detection based on visual features, possibly combined with certain conceptual features, could potentially explain how infants in the learners group made distinct predictions in our experiment without invoking the concepts of ALL and NOT ALL.

One point that is worth mentioning is that there is an interesting parallel between the logical phenomena studied by us and other researchers, in particular, between the notions of exhaustivity, distributivity and sameness on the one hand, and non-exhaustivity, cumulativity and difference on the other hand, which is related to their logical properties and might eventually affect the amount of evidence needed to invoke the underlying concepts. While the former notions are characterized by a 1-to-1 correspondence, i.e. each of the descriptions in (41) is compatible with exactly one scenario in a given context, the latter show a 1-to-many correspondence, i.e. each of the descriptions in (42) may be compatible with multiple scenarios.

(41) **1-to-1 correspondence**

    a. *All agents are pushing a box together.*

    b. *Two cows each are pulling three horses.*

    c. *X is the same as Y.*

(42) **1-to-many correspondence**

    a. *Not all agents are pushing a box together.*

    b. *Two cows, between them, are pulling three horses*

    c. *X is different from Y.*

We speculate that this asymmetry might influence how easily certain concepts are invoked and ultimately may determine how fast certain notions are acquired (in either sense, i.e. either *learned* or *mapped*).

In summary, while the findings reported in the previous chapter provided initial evidence supporting the Mapping Hypothesis, the results presented in this chapter are consistent with multiple interpretations. As a result, we remain agnostic regarding the claims about early conceptual development put forward by the Learning Hypothesis and the Mapping Hypothesis.

Chapter 4

The acquisition of definite descriptions

4.1 Introduction

The previous two chapters explored the relationship between language and thought as well as the question of linguistic and non-linguistic logical primitives. In this chapter we will take a closer look at the connection between language and the world and the challenges children may encounter when acquiring plural expressions. We will focus on the notions of homogeneity and non-maximality and investigate how children manage the potential uncertainty accompanying the interpretation of definite descriptions.

Homogeneity is a phenomenon that has been observed to arise with plural definite descriptions such as *the dogs* or *the mugs*. Plural definites are commonly assumed to refer to the maximal plural entity within the domain specified by the noun, e.g. *the dogs* picks out the largest plurality consisting of individual dogs so that in a situation that involves the four dogs Angel, Bella, Cooper and Daisy, the meaning of *the dogs* will be Angel⊕Bella⊕Cooper⊕Daisy (cf. Link, 1983). However, plural definites exhibit a puzzling behavior in that their meaning seems to depend on the syntactic environment. In affirmative sentences, they behave like universals, whereas under negation they seem to convey an existential meaning (cf. Fodor, 1970; Gajewski, 2005; Križ, 2015, a.o.). Consider for instance the sentences in (43).

78

(43) a. *The mugs are blue.*

　　　b. *The mugs aren't blue.*

While (43a) appears to be only true in a scenario in which all of the relevant mugs are blue and thus corresponds to a universally quantified sentence, (43b) seems to be only true in a scenario where none of the mugs are blue, thus corresponding to an existentially quantified sentence (see Figure 4.1).
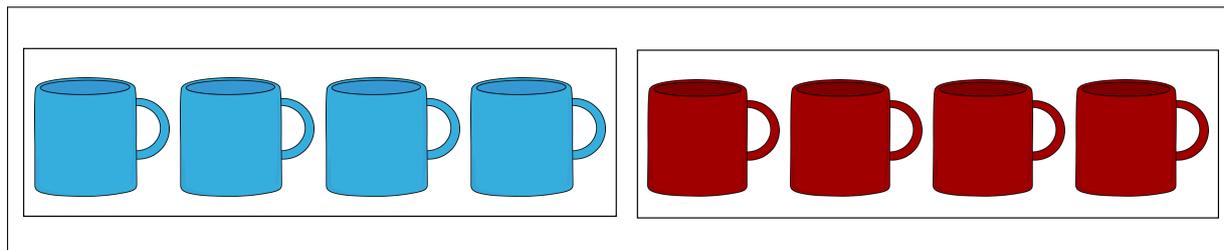
Figure 4.1: Scenario in which *all* mugs are blue (left) and scenario in which *none* of the mugs are blue (right).

Apparently the sentences in (43) do not have complementary truth-conditions and create a gap. The status of situations in which some but not all mugs are blue, e.g. the mixed scenarios in Figure 4.2, is unclear since neither the affirmative sentence in (43a) nor the negated sentence in (43b) seems to be fully acceptable in such cases. This is different when looking at sentences that contain a universal quantifier such as (44a) and (44b).

Figure 4.2: Mixed scenarios in which *some but not all* mugs are blue.

(44) a. *All mugs are blue.*

　　　b. *All mugs aren't blue.*

The universally quantified sentence in (44a), just like the affirmative plural sentence, describes a scenario in which all of the relevant mugs are blue, however, the negative universal sentence (44b), in contrast to the negative plural sentence, is also true in a scenario in which some but

not all mugs are blue, i.e. it is true in the mixed scenarios in Figure 4.2.[1] Sentences that involve an existential quantifier show the opposite pattern: The existentially quantified affirmative sentence in (45a) may be used to describe a 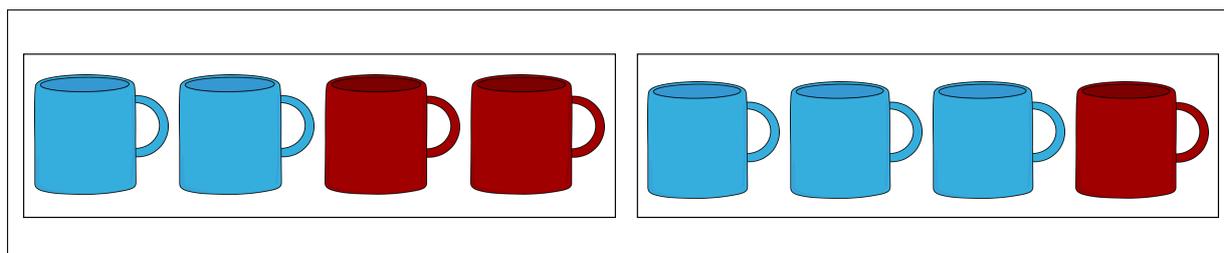mixed scenario where some but not all mugs are blue, while the negative counterpart in (45b), like the negative plural sentence, is true in a scenario in which none of them are.

(45)   a.   *Some mugs are blue.*

       b.   *Some mugs aren't blue.*

Yet, another puzzling fact about plural definites is that they sometimes seem to tolerate exceptions and allow for non-maximal construals. This means that a sentence like (43a) could, under certain circumstances, be an adequate description of a situation in which some but not all mugs are blue, maybe when it is not important or relevant that all objects satisfy the predicate (Brisson, 1998; Lasersohn, 1999; Malamud, 2012; Križ, 2016, a.o.). In this context the example in (46) is discussed by Lasersohn (1999), who argues that the sentence may also be uttered in a situation in which most townspeople are asleep, but some of them are still awake.

(46)   *The townspeople are asleep.*

Intuitively, the number but also the nature of the exceptions matters and the sentence may only be felicitous if the non-sleepers do not have a striking alternative property. For instance, it may be felicitous in a situation in which most of the townspeople are sleeping and a few are quietly reading, but might be less appropriate in a situation in which the non-sleepers are demonstrating in the streets (cf. Brisson, 1998; Lasersohn, 1999; Malamud, 2012; Križ, 2016; Haslinger, 2022). So while homogeneity describes the all-or-nothing effect of plural definites, non-maximality refers to their occasional acceptability in such intermediate situations.

---

[1]Sentences that contain a negation and a quantifier like the examples in (44b) and (45b) may give rise to scope ambiguities, i.e. the negated universally quantified sentence in (44b) has at least two possible interpretations which differ in their respective scope relations, as illustrated below.

(iii)   a.   **Negation wide scope:** $\neg\forall x\,[\text{mug}(x) \rightarrow \text{blue}(x)]$
             'It is not the case that all mugs are blue.'

       b.   **Negation narrow scope:** $\forall x\,[\text{mug}(x) \rightarrow \neg\text{blue}(x)]$
             'All mugs are such that they are not blue.'

We are concerned here with the interpretation in which the negation takes wide scope.

Both phenomena, homogeneity and non-maximality, have also been observed to arise with singular definite descriptions such as *the dog* and *the mug* (cf. Löbner, 2000; Križ, 2015; Paillé, 2022). Consider the singular counterparts of (43) in (47). While the affirmative sentence containing a definite singular in (47a) is apparently only true in a scenario in which the whole mug is blue, the negated sentence in (47b) seems to be only true in a scenario where the whole mug is not blue (Figure 4.3).

(47)   a.   *The mug is blue.*

     b.   *The mug isn't blue.*

Again the status of intermediate situations, i.e. situations in which some but not all parts of the mug are blue, is unclear. Given the discussion above, the acceptability of (47a) and (47b) in mixed scenarios potentially may depend on conceptual and/or contextual factors.
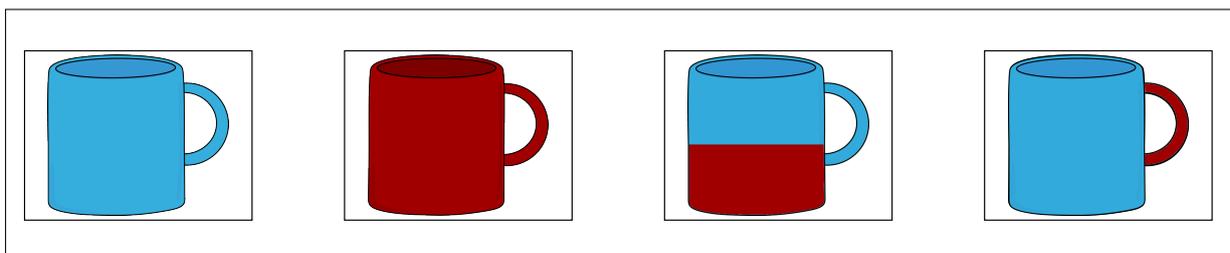
Figure 4.3: Scenario in which *all* parts are blue and scenario where *no* part is blue (two pictures on the left). Mixed scenarios in which *some but not all* parts are blue (two pictures on the right).

4.2   Theoretical background

Following Fodor (1970), many accounts assume that the definite plural carries a presupposition that sentences are true either of all or none of the entities picked out by the noun phrase, otherwise they will be undefined, i.e. neither true nor false (Schwarzschild, 1994; Löbner, 2000; Gajewski, 2005, a.o.). Fodor's main interest was whether definite descriptions are semantically similar to universal descriptions, in particular, whether the definite article exhibits the same semantic behavior as universal quantifiers such as *all* and *every*. The initial observation was that certain contrasts are infelicitous. Consider for instance the difference between the conversations in (48) and (49).

(48) a. *John read some of the books.*

b. *No, John read all of the books.*

(49) a. *John read some of the books.*

b. #*No, John read the books.*

Whereas the sentence containing the universal *all* may be contrastively used to the existentially quantified sentence containing *some* as in (48), the definite article *the* apparently cannot be used in the same way (49). The same point is illustrated by the difference between the following two dialogues:

(50) a. *Has John read all of the books?*

b. *No, some of them.*

(51) a. *Has John read the books?*

b. #*No, some of them.*

The difference between (50) and (51) and the oddness of (51) seems to suggest that the definite in fact does not convey a universal meaning. On the other hand, in sentences such as (52) plural definites do behave like universals since (52) appears to be only true if John read all of the books. However, in negated sentences such as (53) plural definites do not show the same semantic behavior as universals, i.e. (53) may not mean that it is not true that John read all the books, but seems to convey the meaning that John did not read any of the books.

(52) a. *John read the books.*

b. $\approx$ John read all of the books.

(53) a. *John didn't read the books.*

b. $\approx$ John didn't read any of the books.

It seems that in situations where he read some but not all of them such sentences are not straightforwardly interpretable. In sum, plural definites seem to behave like universals in that they require that a predicate holds of all entities that fall under the noun phrase description, but at the same time, they do not stand in the same relation to other quantifiers as universals do. In addition, plural definites exhibit this sort of gappiness when considering intermediate scenarios.

82

Together with the fact that in negated sentences they receive a weaker interpretation it is unclear what quantificational force at all is associated with them. The lack of contrast with other quantifiers and the specific behavior when combined with negation led Fodor (1970) to propose that plural definites do not conform to standard quantifiers but instead carry a homogeneity presupposition which is not satisfied in scenarios that fall between *all* and *none* scenarios. Presuppositions as, informally speaking, pre-conditions for the interpretation of expressions arise in affirmative sentences, but also remain constant under negation. As such, homogeneity is formalized as an integral part of the meaning of definite plurals and expected to occur likewise in affirmative and negated sentences. According to presuppositional accounts, the indeterminacy of definite plurals in intermediate scenarios is treated on a par with the oddness of sentences such as *The king of France is bald* in situations where there is no king of France. In intermediate scenarios sentences containing plural definites are neither true nor false, thus we would expect that it should be difficult or impossible to evaluate them and that speakers would neither fully agree with affirmative sentences nor with negated sentences being true descriptions of such scenarios.

The truth-conditional anomalies of definites have also been discussed in relation to vagueness (e.g. Burnett, 2017; Haslinger, 2022) – the idea that the meanings of certain linguistic expressions are not precisely determined and encompass a truth-conditional gray zone, leading to uncertainty in some contexts. Such uncertainty may occur for instance when it is unclear whether a predicate applies or not, or, more generally, whether a sentence is true or false. The empirical hallmarks of vague expressions are such borderline contexts, in which linguistic judgments may be unclear or conflicting, fuzzy boundaries and that they may give rise to the paradox of the heap (the Sorites paradox) (cf. Burnett and Sutton, 2020). On this view, definite plurals are treated analogously to vague expressions such as *tall*, the paradigmatic example of a vague predicate, which is well known for being context-sensitive and exhibiting borderline cases where it is difficult to determine whether it applies or not (for experimental evidence see Bonini et al., 1999; Alxatib and Pelletier, 2011). These approaches predict that affirmative and negated definite plural sentences will be neither clearly true nor clearly false in intermediate

scenarios.[2]

A very different approach to homogeneity is adopted by implicature-based theories (cf. Magri, 2014; Bar-Lev, 2021). In Magri (2014) for instance, it is assumed that definite plurals have an existential meaning and that the observed homogeneity effects are derived through scalar implicatures, analogously to the strengthened meaning of sentences that contain *some*. Scalar implicature are assumed to arise as a result of comparison with stronger alternatives and may be obtained by the insertion of an exhaustivity operator which functions as a covert *only* in sentences such as (54) (cf. Fox, 2007).[3]

(54)   EXH(*John read some of the books*)

    a.   **Plain meaning:** John read some of the books.

    b.   **1st order implicature:** John read only some, but not all of the books.

This notion is extended to plural definites. Specifically, it is argued that the basic semantic contribution of plural definites is an existential meaning (55a). The universal construal in affirmative sentences is accounted for by an iteratively applied exhaustification operation. Applying the exhaustivity operator to the existential meaning yields the *only-some* implicature in (55b), which then is strengthened again, yielding the *not-only-some* inference in (55c).

(55)   EXH(EXH(*John read the books*))

    a.   **Plain meaning:** John read some of the books.

    b.   **1st order implicature:** John read only some, but not all of the books.

    c.   **2nd order implicature:** John read not only some, but all of the books.

In negative sentences strengthening does not occur since the plain meaning represents already the logically strongest meaning. Therefore, no implicature is triggered and the definite plural is interpreted existentially as in (56a).

---

[2]It is important to note that in Haslinger (2022) it is argued that there may be two types of intermediate contexts, one which comes with clear-cut boundaries, and one which involves an interpretative gray zone. Thus, according to this account, in some intermediate contexts plural definites are predicted to yield crisp judgments.

[3]The rationale behind it is that someone upon hearing (54) will assume that the speaker, given that she is as informative as possible, would have uttered the stronger alternative if she had intended to express a stronger meaning, in this case *John read all of the books*, and thus will infer that this alternative must be false (cf. Grice, 1975).

(56) *John didn't read the books.*

    a.   **Plain meaning:** John didn't read any of the books.

Negated sentences are thus expected to lead to clear judgments in intermediate scenarios, i.e. they are expected to be rejected. judgments on affirmative sentences may be less consistent, depending on whether the implicature is derived or not. If the implicature is calculated, affirmative sentences should be also rejected, but if not, definites will be interpreted existentially (i.e. non-maximally) and should be accepted. Thus, judgments may potentially be asymmetric for affirmative and negated sentences in such scenarios. Magri (2014) assumes that this is context-dependent in that in contexts where the indefinite plural does not trigger a scalar implicature, the implicature of definite plurals will not arise either.[4] The implicature-based account proposed by Bar-Lev (2021) assumes a quite different architecture, particularly in how implicatures are calculated and how context interacts with alternatives. It permits non-maximal construals of affirmative sentences, though to a lesser extent for negated sentences. While it makes roughly the same predictions regarding potential asymmetries between affirmative and negated sentences in intermediate scenarios, in contrast to Magri's account, it attempts also to explain why negated sentences may occasionally be acceptable in such cases.

All of these theories predict that plural definites may lead to interpretative uncertainty in intermediate situations, though for different reasons. Presuppositional accounts attribute this uncertainty or oddness to presupposition failure. Theories that model homogeneity as vagueness assume that it arises because the meaning of definite plurals is inherently indeterminate to some degree. According to implicature-based theories, such oddness may result from a conflict between the literal and the enriched meaning. From a behavioral perspective, distinguishing between these approaches is not always straightforward. However, without further assumptions the first two types of theories predict that uncertain judgments should occur symmetrically, i.e. equally for affirmative and negated sentences, whereas implicature-based accounts suggest an asymmetry, with affirmative sentences potentially creating greater uncertainty than negated

---

[4]Magri (2012, 2014) argues that the same holds for generic plurals in sentences such as *Ravens are black*. In Appendix C, we report the results from a study on the interpretation of conjunctive generic plural sentences in non-maximal and maximal scenarios which suggest that those are equally accepted in both types of scenarios by preschool-aged children.

ones (cf. Križ and Chemla, 2015; Bar-Lev, 2021; see also Spector, 2016).[5]

Homogeneity at the subatomic level has received less attention, though it has often been noted that homogeneity effects likewise arise with singular definites as in (57) and (58) (e.g. Löbner, 2000; Križ, 2015, 2016; Paillé, 2022).

(57) a. *John read the book.*

b. ≈ John read the whole book.

(58) a. *John didn't read the book.*

b. ≈ John did not read any of the book.

In principle the accounts of homogeneity of plural definites can be extended to singular definites and the parthood relation proposed by Link (1983), which holds between atomic individuals and pluralities, can be extended to subatomic parts and individuals (cf. Löbner, 2000; Križ, 2015, 2016; Paillé, 2022). For instance, Löbner (2000) assumes that, as in the case of plurals, such effects are due to a presupposition which leaves sentences containing singular definites undefined in intermediates scenarios, in which not all subparts of an individual object have the property in question. As with plural definites, we would expect singular definite descriptions to give rise to interpretative uncertainty in intermediate scenarios, where a predicate applies neither to all nor to none of an object's subparts. Homogeneity effects should arise symmetrically for affirmative and negated sentences when analyzed in terms of a presupposition or vagueness but may be asymmetric if homogeneity is derived through exhaustification. However, it is also possible that singular and plural definites differ, particularly if homogeneity of the former depends more strongly on conceptual knowledge, as has been claimed more recently (cf. Paillé, 2022).[6]

---

[5]The proposals in Križ (2015, 2016) and Križ and Spector (2021), though each assuming a different architecture, also predict a symmetric pattern, i.e. that positive as well as negative definite plural sentences may come with uncertainty or oddness in intermediate situations, while accounting for their occasional acceptability in such cases in terms of contextual restrictions. In fact the theory in Haslinger (2022) assumes both that plural definites are vague to some extent and that contextual aspects may determine precise interpretations.

[6]Relatedly, predicates may differ in their logical strength (e.g. *dry* vs. *wet*)) (cf. Yoon, 1996; Paillé, 2022) and in how they distribute over an individual's parts (e.g. *white* vs. *intelligent*) (Löbner, 2000).

4.3 Previous findings

While figuring out how to map linguistic expressions to the entities they pick out in the world is generally not a trivial task, it becomes especially challenging when those expressions have indeterminate meanings. How do children acquire such expressions? And how do they navigate the uncertainty that such expressions may create? Little is known about the acquisition process given the limited amount of experimental research on homogeneity. Overall the available experimental data on adults appears to confirm the indeterminacy of plural definites in non-maximal situations noted in the theoretical literature (Schwarz, 2013; Križ and Chemla, 2015; Cremers et al., 2017; Tieu et al., 2019). In contrast, relevant developmental studies have yielded mixed findings regarding how children interpret and acquire definite descriptions, suggesting a divergence between children's and adults' interpretations (e.g. Karmiloff-Smith, 1979; Munn et al., 2006; Caponigro et al., 2012; Simon-Pearson and Syrett, 2018; Tieu et al., 2019; Moyer et al., 2019).

Križ and Chemla (2015) conducted a series of experiments investigating adults' interpretations of plural definites in various linguistic environments, introducing a novel methodological approach to the study of homogeneity. The study employed two different methods to examine how speakers interpret definite plural sentences in intermediate, so-called *gap* scenarios, which fall between *all* and *none* scenarios. In the first part of the study, one group of participants was given the option to judge test sentences as either 'completely true' or 'not completely true', a second group had the option to judge them as either 'completely false' or 'not completely false'. The study involved simple affirmative and negated sentences such as (59a) and (59b), but also more complex ones in which a definite description was embedded under *all* as in (60a), and *no* as in (60b).

(59)   a.   *The triangles are blue.*

   b.   *The triangles aren't blue.*

(60)   a.   *In all the cells, the symbols are blue.*

       b.   *In none of the cells, the symbols are blue.*

In the critical trials, simple sentences were presented together with depictions of arrays of colored geometric shapes which, to varying degrees, were heterogeneous so that some of them had the relevant property, others not (e.g. an array of triangles of which some were blue, others green). The second type of sentences was accompanied by depictions of four arrays, some of which were heterogeneous, others being homogeneous (e.g. two arrays in which all symbols were blue and two in which some were blue, others green). Control trials involved sentences containing *all* such as the examples in (61) and (62).

(61)   a.   *All the triangles are blue.*

       b.   *Not all the triangles are blue.*

(62)   a.   *In all the cells, all the symbols are blue.*

       b.   *In none of the cells, all the symbols are blue.*

Across experiments, the obtained results suggest that in gap scenarios, test sentences were treated differently from control sentences, which yielded overall 'completely true' and 'completely false' judgments (e.g. in Exp. A1 simple positive *all*-sentences received ∼100% 'completely false' and negative ones ∼100% 'completely true' responses). Definite plural sentences on the other hand were mostly judged as 'not completely true' by participants from the first group and 'not completely false' by participants from the second group. Overall the proportion of 'completely true' and 'completely false' responses for plural definites in gap scenarios was quite low and fell well short of totaling 1 (e.g. in Exp. A1 simple positive sentences yielded in sum ∼25% and simple negative sentences ∼50% 'completely true' and 'completely false' judgments).[7] The effect was weaker for negated sentences such as (59b) and not significant for embedded definites under *no* (60b), which both were often judged as 'completely false' (in Exp. A1 negative definite sentences were judged as 'completely false' in ∼50% of the cases; in Exp. A2 definites embedded under *no* yielded ∼75% 'completely false' responses). Together

---

[7]The assumption is that a truth-value gap is indicated when plural definites give rise to different response patterns than *all*-sentences, specifically when definites are judged as neither true nor false in intermediate situations and the acceptance rates for positive and negative sentences do not sum to 1.

these findings can be taken as evidence that indeed definite plural sentences systematically give rise to a truth-conditional gap in intermediate scenarios, while also pointing to a potential asymmetry between negated and affirmative sentences. In the second part of the study, where ternary judgments ('completely false', 'neither', 'completely true') on such and similar stimuli were collected from another group of participants, the pattern of 'completely false' and 'completely true' responses from the previous set of experiments was replicated. However, this time definites, instead of a high rate of 'not completely true' and 'not completely false' judgments, yielded a high proportion of 'neither' responses in gap scenarios (e.g. in Exp. B1 the rate of 'neither' responses for positive definites was ~75%, and for negative ones ~50%), again confirming the indeterminate status of definite plurals in such situations.

Tieu et al. (2019) extended this line of inquiry to children, investigating how preschoolers (French-speaking 4-to-5-year-olds) and adults interpret affirmative and negated sentences containing plural definites in intermediate scenarios. Additionally, existentially quantified with *certains* ('some') were included to find out whether there is a developmental link between scalar implicatures and homogeneous interpretations of plural definites. The first experiment involved a truth-value judgment task in which participants were presented with sentences such as those in (63)–(64) and pictures displaying four objects.

(63) a. *Les ballons sont rouges.*

'The balloons are red.'

b. *Les balles ne sont pas rouges.*

'The balls are not red.'

(64) a. *Certains chapeaux sont roses.*

'Some hats are pink.'

There were three different types of scenarios presented along with sentences, one was an intermediate scenario (gap scenario), in which only two objects had the color indicated in the sentence (e.g. four balloons, of which two were red and two were yellow), one was a universal scenario (*all* scenario), in which all objects had the color (e.g. four red balloons), and one was

89

a scenario where none of the four objects had the specified color (*none* scenario) (e.g. four yellow balloons). The acceptability of definite plural sentences was tested in all three types of scenarios, with critical trials focusing on cases where definite plural sentences were paired with gap scenarios. Indefinites were accompanied by universal scenarios to determine whether participants derived the scalar implicature associated with 'some'.

Adults behaved as expected in that they predominantly accepted positive definite plural sentences in universal scenarios, in which all objects had the property expressed in the sentence, and rejected them in *none* scenarios. Conversely, they rejected negative sentences in *all* scenarios and accepted them in *none* scenarios. In this respect, children were adult-like in their performance. However, in gap contexts, the response patterns diverged: Adults showed some willingness to accept negative definite plural sentences (acceptability rate ∼25%), but overly rejected the positive counterparts (acceptability rate ∼0%). Children on the other hand exhibited a tendency to accept positive sentences (acceptability rate ∼30%), but overly rejected negative ones in such cases (acceptability rate ∼0%). Existentially quantified sentences in universal scenarios yielded comparable acceptability rates in children (46%) and adults (50%). Overall the results suggest that both children and adults showed sensitivity to the homogeneity property of definite plurals, but that children sometimes allowed for non-maximal interpretations of positive sentences, but not of negative ones. Adults seemed to access non-maximal interpretations of negative but not positive sentences. Scalar implicatures with 'some' were equally often derived by both groups.

Participants were further classified based on individual response patterns in gap contexts and scalar implicature trials. The majority of children displayed homogeneous responses, rejecting both positive and negative definite plural sentences in gap contexts. However, only a subset of these children also derived the scalar implicature with 'some'. Another subgroup accepted positive but rejected negative plural definites in gap scenarios, possibly treating the definite as existential. Most children in this subgroup also accepted existentially quantified sentences in universal scenarios, suggesting they did not generate the scalar implicature. Among adults, most participants belonged to the homogeneous subgroup, with some computing the scalar

implicature and others not. Another subgroup rejected positive but accepted negative plural sentences in gap scenarios, seemingly interpreting the definite universally. Most of these participants did not compute the scalar implicature with 'some'. The pattern of individual responses was thus diverse, suggesting that homogeneity with plural definites and scalar implicatures with 'some' are not necessarily related. Additionally, one might speculate that some children, but no adult, interpreted the definite existentially, as indicated by a subgroup of children who accepted positive sentences in gap scenarios while rejecting negative ones. The asymmetry in children's responses is, in principle, consistent with an implicature account of homogeneity. However, further assumptions would be required to explain why some children apparently derived the homogeneity implicature while simultaneously failing to compute the scalar implicature with 'some'.

The second experiment in Tieu et al. (2019) used similar stimuli but introduced a new design in which participants judged sentences by choosing between a minimal, an intermediate and a maximal reward (one vs. two vs. three strawberries). In critical trials where plural definites were presented in gap scenarios, the vast majority of adults gave homogeneous responses, selecting the minimal and intermediate rewards for both positive and negative sentences in almost 100% of the cases. In contrast, only about half of the children followed this pattern. The other group of children selected maximal rewards for positive sentences and minimal rewards for negative sentences, again possibly interpreting the definite existentially. Among the former subgroup, some children, like most adults, derived the scalar implicature with 'some', whereas no child in the latter group did so. As in the first experiment, the pattern suggests that many children and adults recognized the distinctive status of definite plural sentences in gap scenarios. However, the data of some young participants also indicates that there may be a developmental stage at which children have not yet fully acquired the adult-like interpretation of plural definites and instead assign an existential meaning to them, without deriving any homogeneity implicature. Alternatively, one may assume that children who accepted positive and rejected negative sentences were more prone to accept non-maximal construals of regularly maximally interpreted definites, possibly not knowing how context may restrict such meanings, and that their willingness to accept a weaker reading was more nuanced when interpreting positive sentences.

Further evidence for differences between children's and adults' interpretations of plural definites comes from studies that examined only affirmative sentences. For instance, in a series of experiments that involved comprehension as well as production tasks, Karmiloff-Smith (1979) found that 4-and-5-year-old French-speaking children assigned a non-maximal interpretation to definite plurals in affirmative sentences, i.e. understood them as adequate descriptions of intermediate situations, thus pointing to the possibility that preschool-aged children might not yet have acquired maximal interpretations. On the other hand, Munn et al. (2006) report that most English- and Spanish-speaking preschoolers (ages 3 to 5) in their study, which involved an act-out-task, interpreted definite plurals maximally. Conversely, a set of experiments with 4-to-7-year-old children by Caponigro et al. (2012) suggests that maximality does not necessarily underlie children's judgments about definite plural sentences in English before school-age, as it was found that only older children (6- and 7-year olds) were fully adult-like and, like adult participants, assigned maximal interpretations to such sentences. The view that from a certain age speakers derive a strong meaning for the definite plural in positive contexts, but that maximality might be acquired only later in the course of development is further supported by findings on English-speaking children reported in Moyer et al. (2019), which confirmed that preschoolers (between 3 and 5 years) are more willing to accept non-maximal construals of plural definites than adults, and a study by Simon-Pearson and Syrett (2018), in which children of the same age were shown to treat definite plurals on a par with bare plurals.

Another line of research suggests that children and adults also differ in how they handle exceptional or intermediate cases of single objects (Shipley and Shepperson, 1990; Sophian and Kailihiwa, 1998; Brooks et al., 2011; Aravind and Syrett, 2017, a.o.). A seminal study by Shipley and Shepperson (1990) and subsequent research established that children, in contrast to adults, include partial, in other words non-maximal, objects when counting and comparing quantities. In a series of experiments, Shipley and Shepperson (1990) presented English-speaking 3-to-6-year-olds and adults with arrays of objects and requested them to count those. Participants were presented with various types of arrays, for instance one was a homogeneous set of three whole, intact objects and two detached parts of a whole object (e.g. three whole forks and two detached parts forming a whole fork), one was a heterogeneous set and involved whole objects and par-

92

tial objects belonging to different categories, and one was a homogeneous set of whole and partial objects belonging to an unknown nonsense category. The detached parts were always such that they would have formed a complete object and were presented either in an aligned or in an unaligned arrangement. Requests were either specific (e.g. 'Can you count the forks?') or general (e.g. 'Can you count the things?'). Most adults counted each whole object on specific request trials and whole as well as partial objects in general request trials. The majority of children, however, showed a peculiar behavior in that they consistently included each partial object in their counts (e.g. in Exp. 1 in 74% of the general request trials and 72% of the specific request trials). Though overall weaker in older children and when parts were presented aligned, this tendency persisted across conditions, even when children were informed that the two detached parts formed a single complete object, suggesting that they treated partial and whole objects alike and considered both as countable units that fall under the noun phrase description (e.g. *fork*).

An investigation by Brooks et al. (2011) documented that this behavior is not restricted to the counting routine. Adults and young preschoolers (3-and-4-year-olds) were tested on a variety of tasks, which all involved whole as well as broken, partial objects. For instance, participants were asked to compare quantities (e.g. 'Who has more shoes?') when shown two whole objects (e.g. two shoes) and one object broken into three parts (e.g. one shoe cut into three pieces). Another task was meant to elicit responses containing nouns marked for singular and plural and involved either a single whole object (e.g. one fork), two whole objects (e.g. two forks), or a single object broken into two pieces presented in an aligned arrangement (e.g. a single fork sliced into two pieces), whereby participants were asked 'What does X have?'. In further conditions, participants were asked to place a specific object in a circle (e.g. 'Can you put the ball in the red circle?') or to touch every object (e.g. 'Can you touch every plate?') when exposed to single whole objects and broken aligned pieces of an object. Interestingly, across all conditions, children, unlike adults, exhibited the tendency to treat the pieces of objects on a par with whole objects, i.e. children treated them as units of comparison (children: 74.1%; adults: 0.4%), satisfied the request with *every* by touching each partial object (children: 73.3%; adults: 8.3%), and, though to a lesser extent, satisfied the the request with *a* by putting a partial

93

object in the circle (children: 37.5%; adults: 6.7%) and referred to two pieces of an object with a plural noun (children: 30.2%; adults: 0%). Altogether the pattern may be taken as evidence that compared to adults, children are more inclusive when deciding what satisfies a count noun description, and that this interpretation is quite consistently applied across various tasks and linguistic contexts.

However, some findings suggest that children are not blind to the distinction between whole and partial objects and the degraded status of the latter as referents of count nouns. In a study by Syrett and Aravind (2022), which involved a referential task where 3-to-6-year-old children and adults were presented with scenes that featured two characters and asked to provide ternary judgments on the behavior of one of the characters, children exhibited a more nuanced pattern. The task was such that in each trial the first character asked the second character to retrieve an object (e.g. a girl said to a fox 'Please get me a tennis ball') and the second character would always leave and then return with a partial object or a whole object (e.g. a partial tennis ball or a whole tennis ball). For one group the requests were formulated in a neutral way, for the other group those contained an explicit statement about the speaker's goal (as e.g. in 'I want to go outside and play tennis. Please get me a tennis ball.'). Participants could reward the performance of the complying character either with a gold trophy, a blue ribbon or with a black cross.

On control trials, children as well as adults mostly awarded the gold trophy when the character returned with a whole object that matched the noun phrase description (children: 93.2%; adults: 100%), and a black cross when the helper returned with a whole object not matching the noun phrase description (children: 73.9%; adults: 95.6%), whereby children sometimes selected the blue ribbon instead (23.9%). In trials with partial objects, the vast majority of adults gave the intermediate reward ($\sim$85%). Children did not consistently choose the maximal reward in such trials, but showed more variance in their responses, specifically the older children (above 4;7 years) seemed to consider the speaker's goal as they predominantly selected the intermediate reward, not the maximal one, when the helper returned with a partial object and the goal was stated explicitly ($\sim$75%). These results suggest that children are aware of the fact that partial

objects are not optimal referents and that they also seem to be able to take into account the broader communicative context and adjust their decisions accordingly, while at the same time confirming previous findings by showing that especially younger children are more inclined than adults to accept partial objects as referents for count nouns. Interestingly, in two follow-up experiments adults' willingness to do so also varied with context, indicating that partial objects may be 'good enough' referents even for adults in some situations. Similar findings on context-sensitivity have been reported for singular definites such as *the tennis ball* (Aravind and Syrett, 2017) and, more recently, for plural definites as well (Augurzky et al., 2023). Syrett and Aravind (2022, p. 261) hypothesize that there could be a parallel between children's willingness to accept plural definites in non-maximal situations and their tendency to treat partial, i.e. non-maximal, objects as proper referents for singular count nouns in that in both cases an entity that is smaller than the whole is permitted to fall under the extension of the respective expression and children might struggle with fixing an otherwise imprecise or indeterminate meaning in a specific context.

Altogether these studies suggest that task affordances and context may affect to a certain extent the acceptability of non-maximal instances. But when contextual factors are set aside, adults mostly treat plural definite descriptions as neither clearly true nor clearly false in non-maximal situations, implying that their interpretation may involve a truth-value gap. Moreover, adults also seem to be more conservative when interpreting definite and indefinite singular terms in that, per default, they only allow them to pick out whole objects, and do not consider atypical (e.g. partial) instances as proper referents. In contrast, children appear to be much more tolerant, permitting non-maximal cases when adults do not: When interpreting definite plurals, children are inclined to accept those as descriptions for non-maximal scenarios, similarly, they allow partial objects to serve as referents for singular terms. Thus, it seems that children might not yet be aware of the peculiar status of those expressions in such cases.

4.4 Children's interpretation of definite descriptions

With the exception of the study by Tieu et al. (2019), which found that while some children derive homogeneous meanings for plural definites, others do not, there has been no systematic empirical investigation into how children interpret negative and positive sentences in intermediate situations. Moreover, previous studies have not directly compared singular and plural definites. We therefore attempted to study the topic in more detail and shed some new light on the question of how children acquire the meanings of definite descriptions.

We designed a study that included both plural definites (Experiment 4) and singular definites (Experiment 5) in affirmative and negated sentences (e.g. 'Jack found/didn't find the trumpets/the trumpet') presented alongside two types of intermediate scenarios: One that involves non-maximal objects and sets of objects (partial condition), the other featuring maximal objects and sets of objects (mixed condition). In Experiment 4, partial scenarios involved a non-maximal set of objects from a single category (e.g. 2 out of 4 trumpets), whereas in Experiment 5, they involved a non-maximal object (e.g. half a trumpet). This condition was included to test for children's sensitivity to the all-or-nothing effect, i.e. the homogeneity effect, associated with definite descriptions. The mixed condition in Experiment 4 presented a set of four objects, two from one category and two from another (e.g. two trumpets and two brooms), while in Experiment 5, it involved a single object combining features of two distinct categories (e.g. an object with trumpet and broom parts). Although the referents in the mixed condition were maximal, their internal structure was heterogeneous. This contrast allowed us to test whether children exhibit a general preference for homogeneous objects and sets of objects, or whether they show specific sensitivity to the homogeneity requirement of definite descriptions. If children are sensitive to the homogeneity property of definites, we expected them to reject both affirmative and negated sentences in the partial condition, but not in the mixed condition. Conversely, if their judgments are driven by a preference for homogeneity more generally, the opposite pattern might emerge. Furthermore, such a general preference may manifest differently for single objects versus sets of objects, as conceptual knowledge, particularly knowledge about object

96

categories, may play a greater role in the evaluation of individual objects than of sets of objects.

The primary aim of the study was to determine whether children are sensitive to intermediate cases and if so, whether their judgments are adult-like. We sought to investigate to what extent the properties of intermediate cases influence children's judgments and whether judgments for affirmative and negated sentences will be symmetric or asymmetric. We were also interested in potential differences between singular and plural descriptions, in order to draw conclusions not only about the semantic representation of definite descriptions but also about the acquisition process and the underlying mechanisms for deriving the respective interpretations. Accordingly, our study addressed the following research questions:

(a) Do children, like adults, recognize the indeterminate status of definite descriptions in non-maximal scenarios? If so, we would expect them to show a distinct response pattern in partial scenarios compared to scenarios where sentences containing definites are unambiguously true or false.

(b) Are children's judgments symmetric or asymmetric? Symmetric responses would support presuppositional and vagueness accounts of homogeneity, whereas an asymmetric response pattern would conform to implicature-based theories.

(c) Are homogeneity patterns similar for singular and plural definites? Our goal was to study whether children exhibit the same sensitivity to intermediate cases when interpreting singular definites and whether the properties of such cases affect the interpretation of singular and plural definites in the same way.

4.5 Experiment 4

The first experiment addressed the question how children interpret affirmative and negated sentences that contain definite plurals, similar to the examples in (65), in intermediate scenarios. The experiment was conducted with preschool-aged children in German.

(65) a. *Jack hat die Trompeten gefunden.*

　　　'Jack found the trumpets.'

　　b. *Jack hat die Trompeten nicht gefunden.*

　　　'Jack didn't find the trumpets.'

If children behave as adults have been shown to in previous studies, we expected them to produce a distinct response pattern in situations that involve non-maximal sets (partial condition). This sensitivity could in principle yield a symmetric pattern, i.e. equal acceptability rates for affirmative and negated sentences, supporting theories which make symmetric predictions with respect to the availability of non-maximal construals. But it may also be asymmetric, in line with implicature-based account of homogeneity. However, children might also be unaware of the oddness of plural definites in such situations and simply interpret the definite plural in terms of an existential. In this case, positive statements should be accepted in intermediate situations, whereas negative statements should be rejected. Conversely, if children interpret the definite as having a universal meaning, we would expect them to accept negative statements and reject positive ones. Moreover, if children's judgments are driven by a general preference for homogeneity, we might observe lower acceptance rates for sentences presented in the mixed condition, where referents are maximal but internally heterogeneous, compared to clearly true or false scenarios. In principle, both types of intermediate scenarios, i.e. those that involve non-maximality (partial condition) and those that involve internal heterogeneity (mixed condition), could influence children's judgments, potentially revealing different aspects of how they interpret definite plural expressions.

4.5.1 Methods

The experiment was preregistered (https://aspredicted.org/tsp7-5xm2.pdf).

Participants

32 native German-speaking children (range: 4;00 – 6;04 years; $M = 4;11$ years; $SD = 0.76$; 14 females) participated in Experiment 4. Data from one additional child were excluded due to not meeting our inclusion criteria (see section 4.5.2 for details). Test sessions took place in our lab and in the Natural History Museum (NHM) in Vienna. The study received full approval by the ethical commission of CEU PU (PREBO, Psychological Research Board). Informed parental consent was obtained before the experiment.

Design

The study followed a 2 (scenario: partial, mixed) x 2 (polarity: positive, negative) within-subjects design and involved a tablet version of the Truth-Value Judgment Task (TVJT) (Crain and Thornton, 1998) in which participants were asked to judge whether a statement accurately describes a situation. All test sentences were part of a coherent story about a group of pirates. The test sentences were uttered by one of the characters, a parrot called Paula, and presented along with pictures showing another character, a pirate named Jack, and a set of objects. Children could either reward the parrot for an accurate description of the depicted situation with a strawberry or punish her for an inaccurate description with a lemon. Thus, a reward indicated agreement with the test sentence, while a punishment indicated disagreement. Test trials consisted of three phases: In the first phase children were presented with the initial context, the second phase involved a neutral context, which was the same across all trials, and the third phase comprised the final context. Each test trial started with the presentation of a picture displaying an introductory scene which involved the pirate Jack and a set of four objects inside a ship, followed by a picture of a scene in which Jack was carried away from the ship by the sea. Subsequently, children saw a picture of the final scene showing Jack on the beach.

In the partial condition, the four objects in the introductory scene belonged all to the same category (e.g. four trumpets). In the final scene, only two out of the initially displayed four objects were present (e.g. two trumpets), i.e. the set presented in the final scene of the partial condition

PARTIAL CONDITION

MIXED CONDITION

Pos: *"Jack found the trumpets"*
Neg: *"Jack didn't find the trumpets"*

(a) Initial context: Jack is looking for objects inside the ship.

(b) Neutral context: Jack is carried away by the sea.

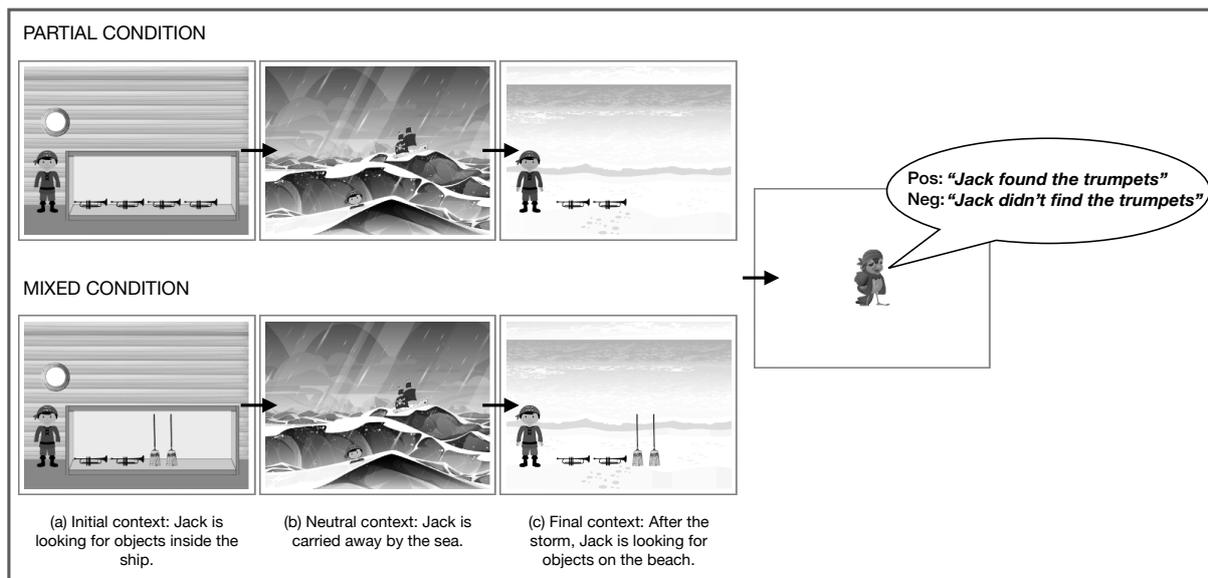(c) Final context: After the storm, Jack is looking for objects on the beach.

Figure 4.4: Test trial structure and stimuli in partial condition (top) and mixed condition (bottom) in Experiment 4

was non-maximal relative to the initially displayed set. In the mixed condition, the four objects shown in the introductory scene belonged to two different categories (e.g. two trumpets and two brooms) and all of them were also present in the final scene, i.e. the set displayed in the final scene of the mixed condition was maximal relative to the set initially shown but heterogenous with respect to its internal structure. Final scenes were presented along with affirmative sentences containing a definite plural (positive condition) and negated sentences with a definite plural (negative condition) (Figure 4.4). Control trials were structurally identical and involved pictures showing sets of four objects of the same category presented together with affirmative and negative sentences. We included four types of control trials. In each control trial the initial context and the final context featured the same set of four objects. In positive true controls, all objects matched the noun phrase description, in positive false controls, none of them did. Similarly, in negative true control trials, none of the displayed objects satisfied the noun phrase description, whereas in negative false control trials all of them did (Figure 4.5). Our dependent variable was the rate of children's yes-responses in the final phase of the test trials.
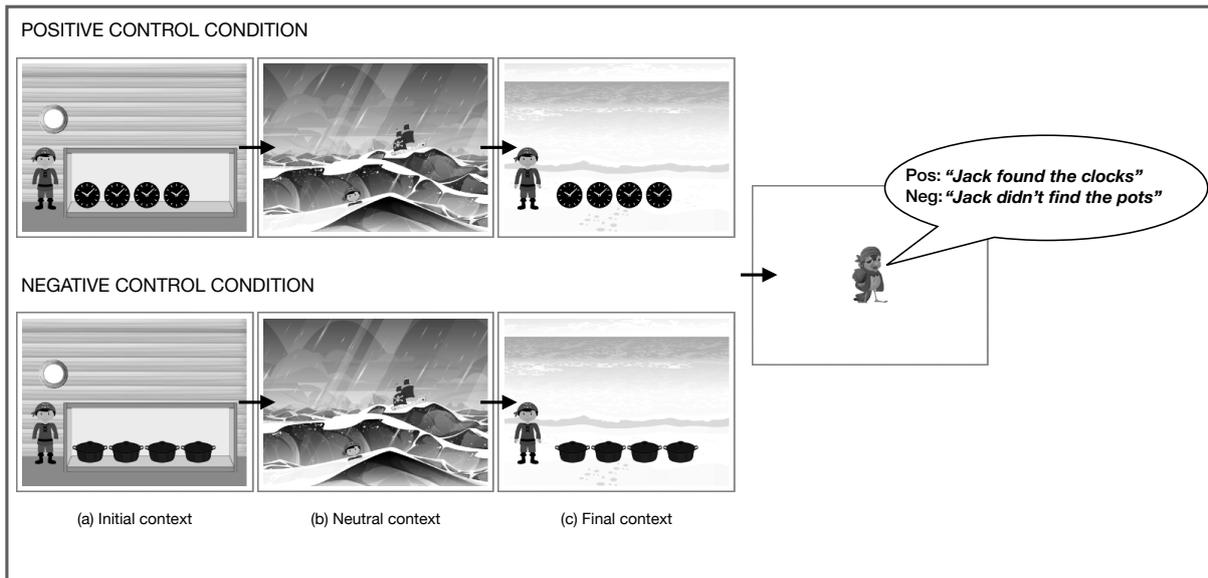
100

Figure 4.5: Control trial structure and stimuli in positive true control trials (top) and negative false control trials (bottom) in Experiment 4

Materials

We used 16 different noun phrases denoting various artifact categories ('hammer', 'knife', 'pair of glasses', 'pair of scissors', 'glass', 'fork', 'whisk', 'brush', 'guitar', 'vacuum cleaner', 'trumpet', 'broom', 'scooter', 'lamp', 'hanger', 'toothbrush') to create our test sentences.[8] Test sentences contained a definite plural in object position and were either affirmative (positive condition) or negated (negative condition) and followed one of the structures below.

(66) **Test sentences**

a. Jack found the X_{PL}.

b. Jack didn't find the X_{PL}.

Each noun appeared in each condition, yielding 32 sentences in total. Another 4 sentences containing different noun phrases ('clock', 'teapot', 'pirate hat', 'pot') were included as control sentences. Half of the control sentences were positive (1 positive true, 1 positive false), half of

---

[8]We did not match the nouns in terms of syllable length, frequency and plural markers, rather our choice was constrained by their compatibility with the plural and singular condition (for considerations guiding the selection of stimuli in the singular condition see section 4.6.1). However, to ensure that the used noun phrases and their referents are known at that age we ran a pilot with a separate group of preschoolers that involved a naming task in which participants were shown pictures of single objects and asked to name those. All children in our sample (n = 10) were able to provide correct descriptions for our items.

them negative (1 negative true, 1 negative false). Control sentences and test sentences followed the same structure (see Appendix B for a complete list of all test and control sentences). All sentences were pre-recorded by a native speaker of Austrian German.

In addition, 64 pictures were created for the test trials. Half of those were used in the partial condition, the other half was presented in the mixed condition. Pictures came in pairs whereby one of them was presented in the initial context and the other one in the final context of a test trial. Each picture showed a set of objects. Objects were placed next to each other and appeared in a canonical position. The size of an individual object depicted on the screen ranged from 0.5 to 5 cm in both length and width. In the partial condition, the pictures presented in the initial context displayed a set of four objects of one category, while the pictures used in the final context showed a set of two objects of this category. In the mixed condition, the scenario depicted on the picture used in the initial phase involved a set of four objects of which two belonged to one category, and the other two to a different category. The picture presented in the final context of the mixed condition showed the same four objects that were present in the initial phase. Throughout the experiment we used the same 8 combinations of object categories displayed in the mixed condition (hammers/knives, pairs of glasses/pairs of scissors, glasses/forks, whisks/brushes, guitars/vacuum cleaners, trumpets/brooms, scooters/lamps, hangers/toothbrushes). Another 4 picture pairs were used in control trials. Pictures shown in the initial phase and the final phase of control trials depicted the same set of four objects belonging to one category. Stimulus presentation followed 16 unique lists, each containing 4 control trials and 8 test trials (2 positive partial, 2 negative partial, 2 positive mixed, 2 negative mixed), whereby no noun was repeated.

Procedure

The experiment was created using PsychoPy 2022.2.4 (Peirce et al., 2019) and hosted by Pavlovia (Open Science Tools Limited, https://pavlovia.org). We presented the experiment on an iPad with a 9.7-inch display and tested participants in a quiet testing room either in the lab or in the museum in the presence of a caretaker. The experiment was introduced as an

adventure game featuring a group of pirates and a talking parrot. The goal of the game was to reward the parrot for accurate descriptions of events with a strawberry, or to punish the parrot for inaccurate descriptions of events with a lemon.

First children were told a short background story which was accompanied by illustrative pictures. Children were introduced to the main characters of the game, the pirates Jack and Pete, Captain Longbeard and the talking parrot Paula, who were sailing with their ship on the sea. The fourth character, the pirate Pete, was included to make a maximal interpretation plausible. The narrative continued with the group being shipwrecked and stranded on a deserted island. Children were informed that the captain ordered Jack to swim back to the sinking ship and to retrieve as many items as possible, and that he ordered Paula to watch Jack and to tell him what is happening. They also learned that Paula was sometimes saying silly things and that she likes strawberries but dislikes lemons. They were then explained that they could reward or punish her for accurate and inaccurate descriptions by touching an image of a strawberry or a lemon.

After 4 practice trials, 8 test trials and 4 control trials were presented in pseudo-random order. In the initial phase of each trial, participants were told that Jack swam back to the ship and went inside, while a picture showing Jack and a shelf containing four objects inside the ship was displayed on the screen. Before he could retrieve anything from the ship, a storm began, and the sea washed him back to land. This served as the neutral context, consistent across all trials, where children viewed an image of Jack being carried away by the sea. Next, a picture depicted Jack alongside a set of objects on the beach, and children were told that, after the storm had passed, he was looking for items that had been washed ashore. They were also informed that Paula was observing Jack. While this image remained on the screen, a pre-recorded sentence was played. The length of sentences varied between 2 and 3 sec. Each trial took 5 sec, after which a picture of a strawberry next to a picture of a lemon appeared on the screen and children were invited to indicate their judgment by touching one of the pictures. There was no time limit for making a decision and the pictures were on screen until participants responded. During the initial practice phase children received feedback on their decisions, while in the test phase no feedback was given. Each participant was compensated with a small toy or a sticker.

4.5.2 Results

We excluded participants who provided correct responses in less than 75% of control trials ($n = 1$). In the remaining sample the overall accuracy rate in control trials was 84.38 % (SE = 2.17). In positive true controls accuracy was at ceiling with children giving correct responses in 100% of the trials. In positive false controls the accuracy was at 96.88% (SE = 3.13). In negative true control trials children were correct in 46.88% (SE = 8.96) of the trials, while in negative false control trials the accuracy rate was at 93.75% (SE = 4.35).

Acceptance rate at test varied considerably across trial types: In the partial condition, children accepted positive sentences in 54.69% (SE = 8.51) of the cases, while negative sentences were accepted in 15.63% (SE = 5.70) of the trials (Figure 4.6). In the mixed condition, the rate of yes-responses for positive sentences was 93.75% (SE = 3.72), compared to only 3.13% (SE = 3.13) for negative sentences (Figure 4.7). When looking at individual response patterns, we found that 14 participants exhibited symmetric responses in the partial condition, with most of them ($n = 11$) consistently rejecting both positive and negative sentences across all test trials. The remaining 18 children displayed an asymmetric response pattern, with the majority ($n = 12$) consistently accepting every positive sentence while rejecting every negative sentence. Only one child showed the opposite pattern (Table 4.1). In the mixed condition, the vast majority of participants ($n = 28$) accepted all positive sentences and rejected all negative sentences across trials (Table 4.2). All 12 participants who consistently accepted positive and rejected negative sentences in the partial condition exhibited the same pattern in the mixed condition. Additionally, 12 of the 14 children who provided symmetric responses in the partial condition displayed the asymmetric pattern in the mixed condition (see also Figure B.1 and Figure B.2 in Appendix B).[9]

Further data analysis was performed with logistic mixed-effects models (Jaeger, 2008) in R

---

[9]To identify outliers, we also determined how many items fell more than two standard deviations away from the overall mean within each polarity. In partial scenarios, there was only one item which exceeded this threshold in the negative polarity condition (ID 11). In mixed scenarios, two items were more than two standard deviations away from the overall mean within the negative condition (ID 8, ID 16), and one item within the positive condition (ID 13).
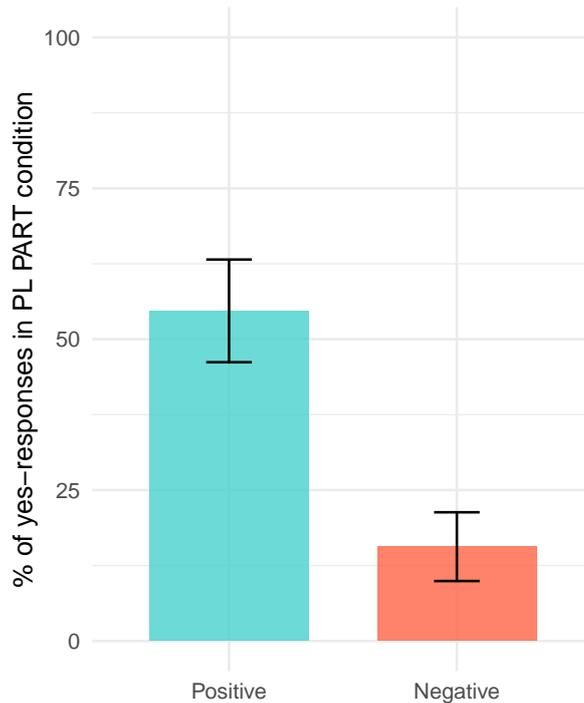
Figure 4.6: **Overall means in partial condition in Exp. 4** Mean percentage of yes-responses provided in positive and negative test trials in the partial condition in Experiment 4. Error bars represent the standard error.

(v4.0.5, R Core Team, 2021) using the lme4 package (Bates et al., 2015). The coefficients of the fixed effects in the mixed-effects models are provided as odds ratios (ORs) with 95% confidence intervals (CIs) where relevant. To perform our main comparisons, we fit a model with Polarity (2 levels: positive, negative), Scenario (2 levels: partial, mixed) and their interaction as fixed effects and included random intercepts for Participants.[10] We conducted model comparison by removing the relevant fixed effect and used χ^2 likelihood ratio tests to obtain p-values.

Main comparisons The analysis revealed that Scenario had a significant effect on acceptance rate ($\chi^2(1) = 10.51$, $p = .001$). We also found a significant effect of Polarity ($\chi^2(1) = 141.26$, $p < .001$). Moreover, the interaction between Polarity and Scenario was significant ($\chi^2(1) = 37.08$, $p < .001$). The effect of positive sentences differed across scenario types in that for mixed sets, relative to the baseline (= negative mixed condition) the likelihood of a yes-response was

---

[10]The initial model included additionally random intercepts for Items, but failed to converge since Item variance was close to 0. The results did not change when we included random intercepts for Items and re-ran the analysis with the optimizer bobyqa. A model with random intercepts for both Participants and Items did not improve fit compared to a simpler model (AIC = 211.64 vs. 209.64; BIC = 229.37 vs. 223.82), and was therefore not retained.
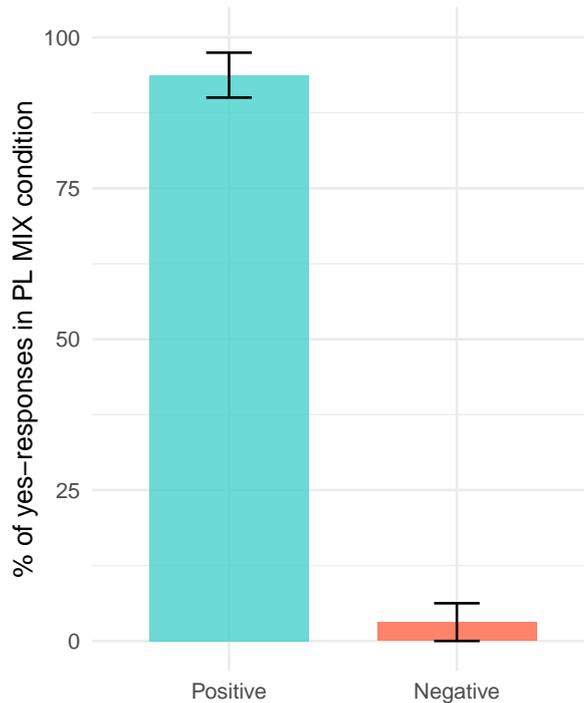
Figure 4.7: **Overall means in mixed condition in Exp. 4** Mean percentage of yes-responses given in positive and negative test trials in the mixed condition in Experiment 4, with error bars showing the standard error.

substantially increased (OR = 90.16, 95% CI $[30.93, 366.37]$, $p < .001$), while this effect was less pronounced and not statistically significant for partial sets (OR = 1.35, 95% CI $[0.71, 2.57]$, $p = .372$).

Developmental trajectory We additionally tested whether participant's Age (centered, in months) had an effect on responses, which, however, was not confirmed ($\chi^2(1) = 0.27$, $p = .606$). There was also no interaction between Age and Scenario ($\chi^2(1) = 0.80$, $p = .372$), or Age and Polarity ($\chi^2(1) = 2.11$, $p = .146$).

Comparison against chance For testing against chance we fit a model that included Condition (4 levels: positive partial, positive mixed, negative partial, negative mixed) as fixed effect and random intercepts for Participants. For each condition, a model with that condition as reference level was fit and it was tested whether the intercept differed from chance level. Our analysis revealed that in the positive mixed condition the likelihood of a yes-response was significantly above chance (OR = 90.04, 95% CI $[12.97, 624.94]$, $p < .001$), while in the positive partial

condition responses were at chance (OR = 1.35, 95% CI $[0.46, 3.92]$, p = .581). Responses in the negative mixed condition were significantly lower than chance (OR = 0.005, 95% CI $[0.0004, 0.05]$, $p < .001$), as were responses in the negative partial condition (OR = 0.05, 95% CI $[0.01, 0.22]$, $p < .001$).

Table 4.1: Participant means in partial condition in Exp. 4

n	Positive (%)	Negative (%)
11	0	0
2	50	0
12	100	0
1	0	50
1	50	50
2	100	50
1	0	100
2	100	100

Table 4.2: Participant means in mixed condition in Exp. 4

n	Positive (%)	Negative (%)
1	0	0
2	50	0
28	100	0
1	100	100

4.5.3 Discussion

Overall the findings suggest that children were sensitive to the peculiar status of plural definites in intermediate situations. Participants exhibited a distinct response pattern when presented with partial, i.e. non-maximal, sets of objects compared to their judgments in maximal scenarios. In the partial condition, acceptance rate for positive sentences was at chance level (54.69%), whereas the rate for negative sentences was below chance (15.63%), although it did not reach zero. In contrast, in the mixed condition, where children were presented with maximal yet internally heterogenous sets, positive sentences were predominantly accepted (93.75%), while negative sentences were overwhelmingly rejected (3.13%). Scenario type had a significant effect on children's choices. Similarly, the pattern in the partial condition differed from

the one in control trials, which involved clearly true and clearly false cases. In control trials in which a positive sentence was paired with a scenario that featured a maximal set of objects of which all matched the noun phrase description, children's acceptance rate was 100%, while negative sentences were almost always rejected in such scenarios (6.25%).

What we also observe is an asymmetric distribution of responses, with positive sentences being accepted more frequently than negative ones, a pattern consistent with implicature-based accounts of maximality (cf. Magri, 2014; Bar-Lev, 2021). According to this view, plural definites have a basic existential meaning. Thus, one possible explanation of the data is that children interpreted plural definites existentially and therefore (mostly) rejected negative sentences in partial scenarios, while occasionally accepting their positive counterparts in such cases because the implicature that would otherwise yield the strengthened maximal interpretation was not computed. Without additional assumptions, such an asymmetry is not predicted by other accounts. Recall that theories explaining homogeneity in terms of a presupposition or vagueness assume that positive and negative sentences should be equally odd in intermediate situations. This finding aligns with previous research on homogeneity in children, in fact the response pattern observed in the partial condition closely mirrors the asymmetric pattern found for non-maximal scenarios in Tieu et al. (2019).

Notably, our results reveal yet another parallel to those reported in Tieu et al. (2019). When analyzing individual response patterns, two major subgroups emerged: One that rejected both positive and negative sentences in partial scenarios, exhibiting a homogeneous response pattern, and another that accepted positive but rejected negative sentences, consistent with an existential interpretation. This finding almost exactly replicates the results from Tieu et al. (2019), where a similar split was observed in their sample. Although each participant provided only two observations per condition, making definitive conclusions premature, this response distribution may suggest that some children had not yet recognized the peculiar status of non-maximal construals. This could be due to them not yet having acquired the implicature associated with plural definites. Alternatively, it is conceivable that these children have accommodated a different context in positive trials in which a non-maximal construal of an otherwise maximal interpre-

tation was felicitous. In any case, the fact that most children, namely 19 out of 32, showed a pattern that is neither consistent with an existential nor a universal interpretation may be taken as evidence that indeed the majority of participants were aware of the particularities associated with plural definites.

4.6   Experiment 5

In Experiment 4, we found that many children exhibited an adult-like response pattern when interpreting plural definites in non-maximal scenarios, rejecting both positive and negative sentences equally. This suggests that these children were aware of the truth-conditional uncertainty in such cases. Another group of children was more permissive, accepting positive sentences, apparently treating the definite as an existential. At the same time, the overall pattern remained asymmetric, which together provides initial support for implicature-based theories. In Experiment 5, we aimed to investigate whether the same pattern would emerge when children interpreted singular definites in two types of scenarios, one that involved partial, i.e. non-maximal, objects and another that involved mixed, i.e. maximal but internally heterogeneous, objects. The primary goal of the experiment was to determine whether definite singular sentences, similar to the examples in (67), elicit the same linguistic judgments as definite plural sentences under comparable conditions.

(67)   a.   *Jack hat die Trompete gefunden.*

'Jack found the trumpet.'

b.   *Jack hat die Trompete nicht gefunden.*

'Jack didn't find the trumpet.'

In principle, if children of a certain age had acquired homogeneity with one type of expression, we would expect them to demonstrate competence with the other as well. However, there were two reasons to suspect potential differences between homogeneity in singular and plural definites. First, prior research has documented children's persistent willingness to include

partial objects across various tasks and linguistic contexts (e.g. Shipley and Shepperson, 1990; Sophian and Kailihiwa, 1998; Brooks et al., 2011; Aravind et al., 2017), which may potentially influence their intuitions about what qualifies as maximal, and second, the objects used in the mixed condition may conflict with children's knowledge about object categories, potentially leading to a response pattern that deviates from that observed in the plural mixed condition.

4.6.1 Methods

The experiment was preregistered (https://aspredicted.org/tsp7-5xm2.pdf).

Participants

32 native German-speaking children participated in the Singular condition (range: 3;08 – 6;05 years; M = 5;03 years; SD = 0.75; 12 females). One additional child was tested but had to be excluded due to not meeting our inclusion criteria. We tested children in our lab and in the Natural History Museum (NHM) in Vienna. The study received full approval by the ethical commission of CEU PU (PREBO, Psychological Research Board) and informed parental consent was obtained before the experiment.

Design

The experiment followed a similar design as Experiment 4 and involved 2 (scenario: partial, mixed) x 2 (polarity: positive, negative) within-subjects conditions and a tablet version of the Truth-Value Judgment Task, but this time test sentences contained a singular definite description. We used the same background story about a group of pirates to embed our test sentences and children were asked to reward the character who uttered the sentences for an accurate description, or to punish it for an inaccurate description. As in Experiment 4, test trials consisted of three phases: The first phase involved the initial context, followed by the middle phase in which a neutral context was presented, and the third phase, which involved the final context.

During the first phase, children saw a picture showing the introductory scene which featured the pirate Jack and a single object inside a ship. Subsequently, in the middle phase, children were presented with a picture that displayed Jack how he was carried away from the ship by the sea. This was followed by the final phase which involved a picture showing Jack on the beach.

In the partial condition, the object shown in the introductory scene was a whole item displaying features of a single category (e.g. a trumpet), whereas in the final scene, the presented object was only a partial one (e.g. half a trumpet), i.e. the object was non-maximal relative to the object initially displayed but homogeneous with respect to its categorical features. In the mixed condition, the object shown in the introductory scene was such that one part exhibited features of one category (e.g. a trumpet), and the other part had features of another category (e.g. a broom). This object was also displayed in the final scene. Thus, the object shown in the final scene was maximal relative to the initially displayed object but heterogeneous with respect to its categorical features. During the final scene children listened to affirmative sentences containing a singular definite (positive condition) and negated sentences containing a singular definite (negative condition) (Figure 4.8). We included control trials that were structurally identical and involved pictures showing whole objects presented along with affirmative and negative sentences (Figure 4.9). There were four types of control trials. In positive true control trials, children saw a picture of a whole object that satisfied the noun phrase description, whereas in positive false controls a whole object was displayed that did not match the description. Negative true control trials involved a whole object that matched the noun phrase description, while negative false control trials involved a whole object which did not fall under the noun phrase description. We measured again the rate of children's yes-responses in the final phase of the test trials.

Materials

The same 16 noun phrases denoting various object categories as in Experiment 4 were used to to create the test sentences. The criteria on which we based our selection of nouns and corresponding objects were: (i) The objects needed to have distinctive visual features that enable

PARTIAL CONDITION

MIXED CONDITION

Pos: *"Jack found the trumpet"*
Neg: *"Jack didn't find the trumpet"*

(a) Initial context: Jack is looking for objects inside the ship.

(b) Neutral context: Jack is carried away by the sea.

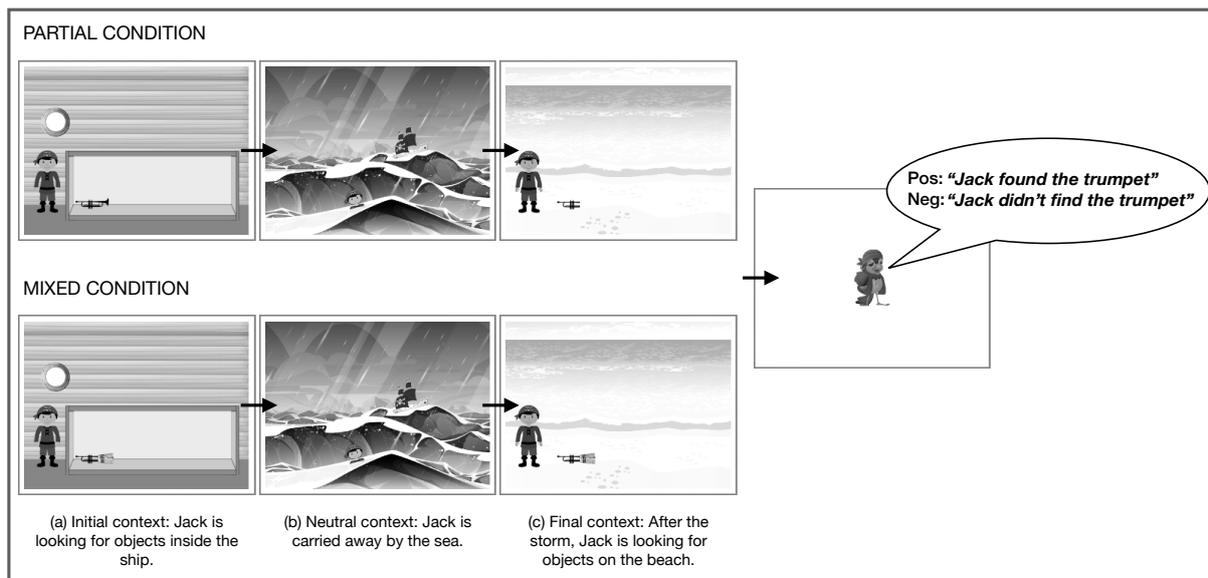(c) Final context: After the storm, Jack is looking for objects on the beach.

Figure 4.8: Test trial structure and stimuli in partial condition (top) and mixed condition (bottom) in Experiment 5

quick identification of their category, (ii) they had to be still recognizable when appearing in the partial condition, (iii) the combination of parts from one object with parts from another object belonging to a different category had to be feasible, and (iv) the objects and their descriptions had to be familiar to preschoolers. All test sentences contained a definite singular noun phrase in object position and were either affirmative (positive condition) or negated (negative condition).

(68) **Test sentences**

    a.   Jack found the X_{SG}.

    b.   Jack didn't find the X_{SG}.

We had a total of 32 test sentences and 4 additional control sentences, which this time also involved singular definites (see Appendix B for a complete list of all test and control sentences used in Experiment 5). All sentences were pre-recorded by a native speaker of Austrian German.

64 pictures were created for the test trials, of which half was used in the partial, and half in the mixed condition. In the partial condition, the picture accompanying the initial context showed a single, whole object. The picture in the final context displayed the partial version of that object.
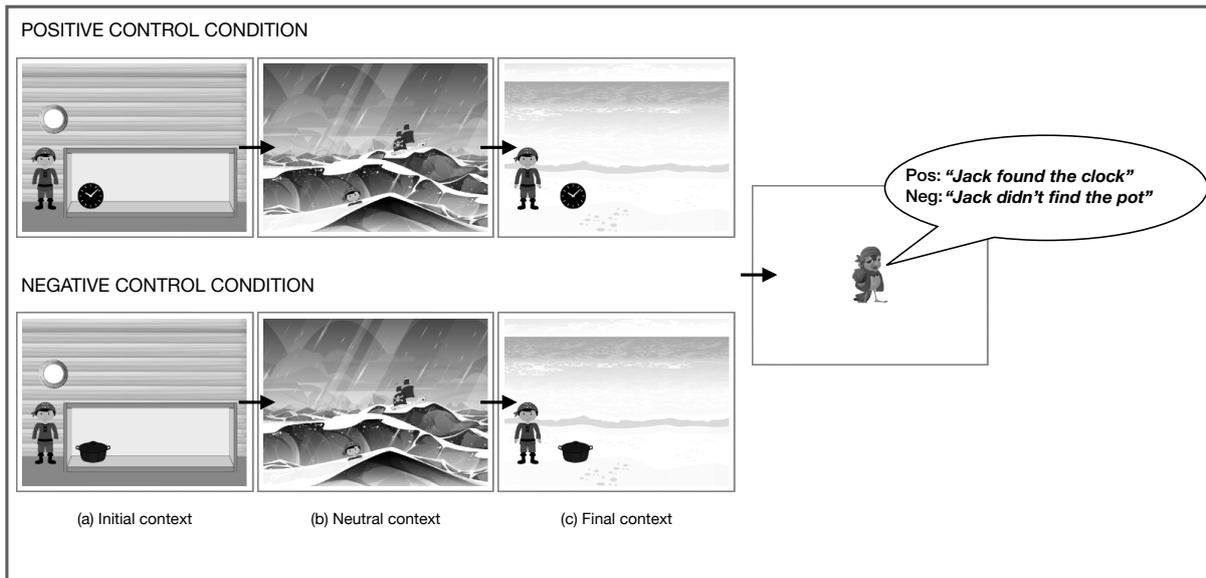
Figure 4.9: Control trial structure and stimuli in positive true control trials (top) and negative false control trials (bottom) in Experiment 5

In the mixed condition, the initial and the final phase involved a picture of an object of which one part belonged to one category (e.g. a trumpet), and the other part belonged to a different category (e.g. a broom) (Figure 4.10). In control trials children saw images of single, whole objects. The depicted objects ranged from 0.5 to 5 cm in both length and width. Each child was presented with 4 control trials (1 positive true, 1 positive false, 1 negative true, 1 negative false) and 8 test trials (2 positive partial, 2 negative partial, 2 positive mixed, 2 negative mixed) in pseudo-random order.

Procedure

The procedure was the same as in Experiment 4.

4.6.2 Results

All participants who did not provide correct responses in at least 75% of control trials were excluded ($n = 1$). Children in the final sample showed overall a high accuracy rate and gave correct responses in 91.41% (SE 2.13) of control trials. In positive true controls accuracy
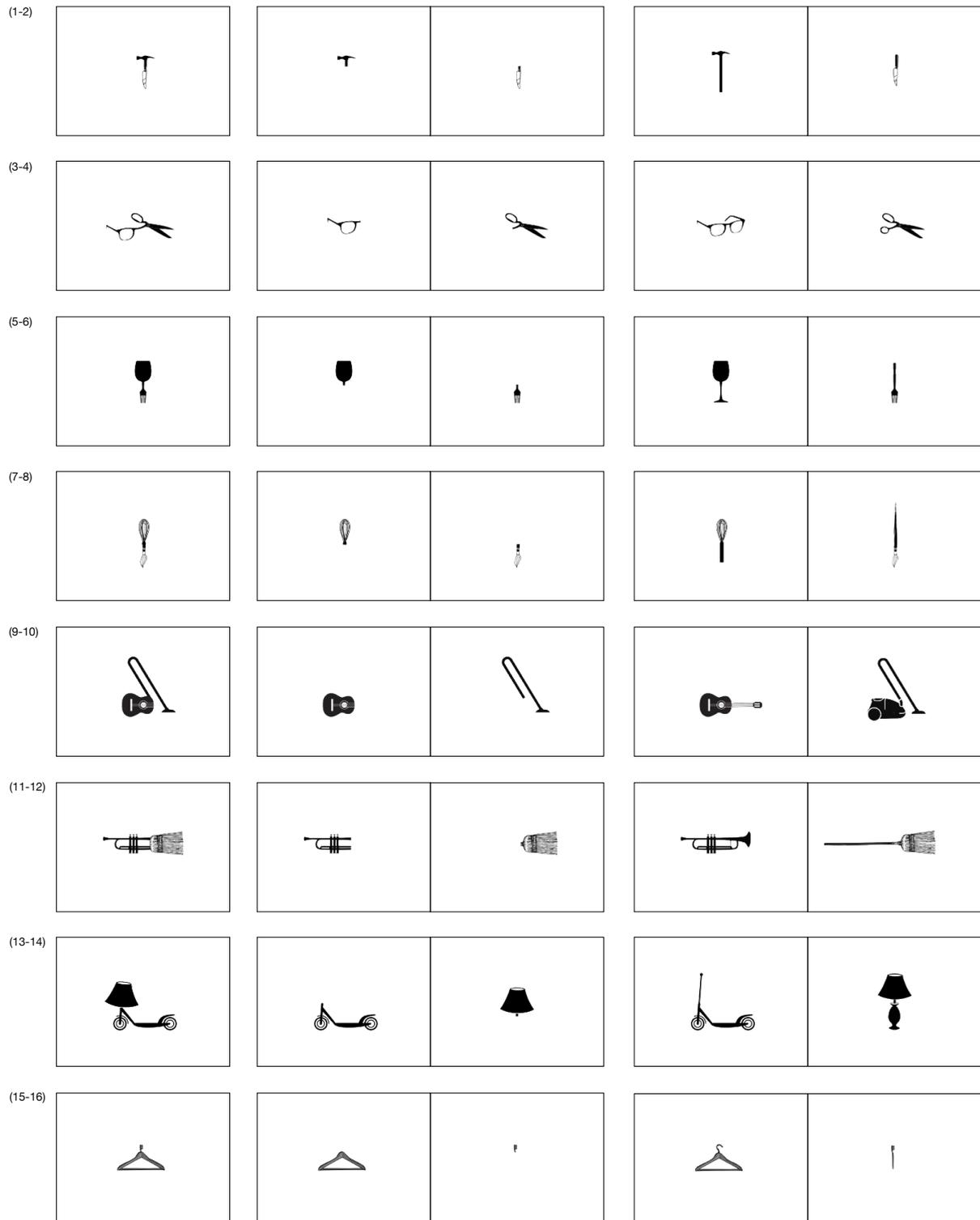
113

Figure 4.10: **Objects in Exp. 5** Mixed (left), partial (middle) and whole objects (right) used in Experiment 5 (ID 1 to 16)

114

was again at ceiling with children providing correct responses in 100% of the trials, just as in positive false control trials (100%). Children were correct in 68.75% (SE = 8.32) of the cases in negative true controls, and gave correct responses in 96.88% (SE = 3.13) of the negative false control trials.

In the positive partial condition, children accepted the test sentences in 87.50% (SE = 4.49) of the cases. In the negative partial condition, the acceptance rate was at 14.06% (SE = 4.62) (Figure 4.11). A different pattern emerged in mixed scenarios: In positive mixed trials, the rate of yes-responses was 48.44% (SE = 7.27), and in negative mixed trials children's acceptance rate was 40.63% (SE = 6.90) (Figure 4.12). When examining individual response patterns, we found that only six children exhibited symmetric responses in the partial condition, whereas 26 of the 32 children displayed asymmetric patterns. The majority of these children ($n = 21$) consistently accepted all positive sentences and rejected all negative ones, with no child exhibiting the opposite pattern. In the mixed condition, eight children demonstrated a symmetric pattern, with five children consistently rejecting both positive and negative sentences. The remaining 24 children showed an asymmetric pattern. However, no clear trend emerged in this condition, three of the eight children who gave symmetric responses and most of the children who gave asymmetric responses ($n = 17$) were inconsistent in their judgments (see also Figure B.3 and Figure B.4 in Appendix B).[11][12]

For our main analysis, we applied the same statistical methods as in Experiment 4. We fit a model with Polarity (2 levels: positive, negative), Scenario (2 levels: partial, mixed) and their interaction as fixed effects and included random intercepts for Participants. The significance of the fixed effects was assessed by removing the fixed effect and comparing the full model to a

---

[11]Additional explorative analyses of individual patterns revealed no meaningful correlation between children's responses in the partial and their judgments in the mixed condition. For instance, not all children who consistently accepted mixed objects as referents also accepted partial objects (though three of the four children did). Similarly, not all children who provided symmetric responses in the partial condition did so in the mixed condition (three out of six did). Six out of the eight children who showed some inconsistency in the partial condition, i.e. who did not always choose the same response for positive or negative sentences when presented with a partial object, also exhibited inconsistency in the mixed condition.

[12]When examining item-level variance, we found that in partial scenarios one item (ID 7) in the positive condition and two items (ID 3, ID 5) in the negative condition exceeded the two-standard-deviation threshold from the overall mean. In the mixed condition, only one item (ID 13) in the positive condition exceeded this threshold.
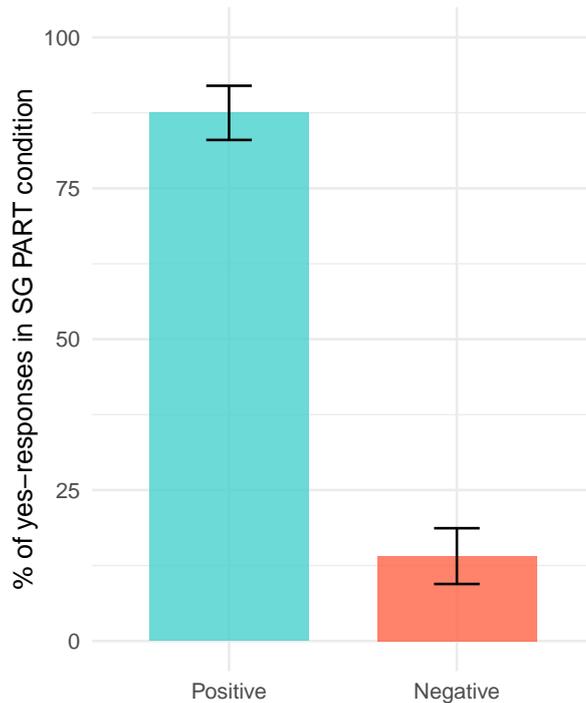
Figure 4.11: **Overall means in partial condition in Exp. 5** Mean percentage of yes-responses provided in positive and negative test trials in the partial condition in Experiment 5, with error bars representing the standard error.

simpler one using χ^2 likelihood ratio tests.

Main comparisons The analysis showed that the effect of Polarity was significant ($\chi^2(1) = 43.81$, $p < .001$). The effect of Scenario was not significant ($\chi^2(1) = 1.20$, $p = .273$), but the interaction between Polarity and Scenario was ($\chi^2(1) = 34.19$, $p < .001$). For partial objects the likelihood of a yes-response was significantly increased in positive trials (OR = 31.16, 95% CI $[9.03, 107.54]$, $p < .001$), while this was not the case when the object was mixed (OR = 1.37, 95% CI $[0.68, 2.76]$, $p = .374$).

Developmental trajectory A further analysis showed that Age (centered, in months) did not have a significant effect on the likelihood of yes-responses ($\chi^2(1) = 0.04$, $p = .850$). There was also no interaction with Scenario ($\chi^2(1) = 1.99$, $p = .158$), or Polarity ($\chi^2(1) = 0.12$, $p = .724$).

Comparison against chance For testing against chance we fit a model that included Condition (4 levels: positive partial, positive mixed, negative partial, negative mixed) as fixed effect and random intercepts for Participants. For each condition, a model was fit with that condition as
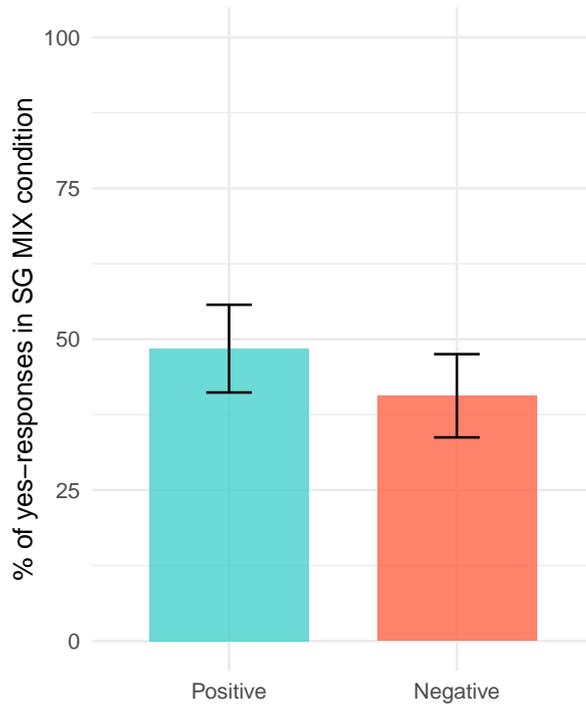
Figure 4.12: **Overall means in mixed condition in Exp. 5** Mean percentage of yes-responses provided in positive and negative test trials in the mixed condition in Experiment 5. Error bars represent the standard error.

reference level. We tested whether the intercept differed from chance level. In the positive partial condition, children were more likely to give a yes-response than expected by chance (OR $= 6.70$, 95% CI $[3.34, 14.68]$, $p < .001$), whereas responses in the negative partial condition were below chance (OR $= 0.16$, 95% CI $[0.08, 0.33]$, $p < .001$). As for the mixed conditions, responses did not differ from chance neither in the positive mixed condition (OR $= 0.94$, 95% CI $[0.58, 1.53]$, $p = .803$), nor in the negative mixed condition (OR $= 0.68$, 95% CI $[0.42, 1.13]$, $p = .136$).

Table 4.3: Participant means in partial condition in Exp. 5

n	Positive (%)	Negative (%)
1	0	0
2	50	0
21	100	0
4	50	50
3	100	50
1	100	100

Table 4.4: Participant means in mixed condition in Exp. 5

n	Positive (%)	Negative (%)
5	0	0
4	50	0
4	100	0
3	0	50
3	50	50
6	100	50
3	0	100
4	50	100

4.6.3 Discussion

Our findings suggest that children were receptive to the experimental manipulations we introduced. In the partial condition, children frequently accepted positive sentences (87.50%), but rarely accepted negative sentences (14.06%). This response pattern resembles that of the control condition, where participants consistently agreed with true positive statements (100%) but almost never with false negative statements (3.12%), although responses showed greater variation when children were presented with partial objects compared to whole objects. The acceptance rate for positive sentences in the partial condition was above chance, while acceptance of negative sentences was below chance. In contrast, in the mixed condition, children accepted positive sentences in 48.44% of the cases and negative sentences in 40.63% of the trials, both of which were at chance level, suggesting greater uncertainty in evaluating these scenarios.

The results from the partial condition indicate that most children did not derive homogeneous interpretations of definite singular sentences, but instead consistently interpreted them in terms of an existential. Children treated sentence such as 'Jack found the trumpet' as felicitous even when Jack found only some parts of the trumpet, rather than requiring him to have found all parts. Similarly, they rejected negative sentences such as 'Jack didn't find the trumpet' when Jack found at least some parts, consistent with an existential meaning. This asymmetric pattern in the partial condition provides prima face support for implicature-based accounts of

homogeneity. However, given previous findings that in contrast to adults children often extend singular count nouns to partial objects across different linguistic contexts, an alternative interpretation is that children understood the all-or-nothing effect of definite singular sentences but differed in what they considered as maximal, possibly shaped by contextual accommodation. Under this view, children treated partial objects (almost) on par with whole objects, leading to overall agreement with positive sentences and rejection of negative ones.

The mixed condition suggests a more complex picture. When presented with a maximal but heterogeneous object (e.g. an object composed of both trumpet and broom parts), children did neither consider the positive nor the negative sentences as clearly true of false, as reflected in their chance-level performance and inconsistent judgments. Crucially, while a scenario in which Jack found only some trumpet parts rendered the positive sentence clearly true, a scenario in which he found some trumpet parts but also broom parts did not lead to clear judgments. This suggests that children did not simply treat any object with some trumpet parts as a valid referent but instead took the remaining parts into account when interpreting the sentence.

Here a key contrast emerged between children's judgments on plural vs. singular sentences. In the plural condition, additional objects in a set did not affect judgments. However, in the singular condition, additional parts of an object clearly influenced judgments. This suggests that while plural definites rely on the part-whole structure of sets, singular definites require additional reasoning about the applicability of the noun phrase (e.g. 'trumpet'). Thus, although the asymmetric pattern observed in both groups seems consistent with implicature-based approaches, suggesting that children assigned an existential meaning to both plural and singular definites (e.g. 'Jack found some of the trumpets'/'Jack found some of the trumpet parts') which eventually got strengthened, the symmetric response pattern in the singular mixed condition raises some questions. The uncertainty in this condition suggests that knowledge about object categories plays a role in the interpretation of singular definites but not in the same way for plural definites. The pattern suggests that vagueness stemming from the noun phrase may have contributed to this uncertainty, potentially influencing judgments in the partial condition as well. However, we cannot fully estimate the extent to which noun phrase indeterminacy

119

affected responses in that condition.

Our study thus demonstrates that homogeneity in plural and singular definites shares similarities but also notable differences. First, in the partial singular condition, the majority of participants accepted non-maximal single objects as referents. In the plural condition, about half of the children accepted non-maximal sets, while the other half rejected positive as well as negative sentences, suggesting that homogeneous interpretations of plural definites were more frequently derived than those of singular definites in children of the same age. Second, in the mixed plural condition, when Jack found all available trumpets plus some additional objects, positive plural sentences were considered acceptable. In contrast, in the singular mixed condition, when Jack found a maximal object composed of all available trumpet parts plus some additional non-trumpet parts, children rather treated these as borderline cases, suggesting that evaluating the part-whole structure of single objects involves additional knowledge compared to evaluating sets of objects. Altogether these findings show that while homogeneity in plural definites is more systematic, homogeneity in singular definites is likely additionally influenced by conceptual knowledge.

An interesting question for future research is whether additional factors, not manipulated in our study, would affect acceptability. For one, contextual affordances might modulate responses, as suggested by previous results (cf. Syrett and Aravind, 2022; Augurzky et al., 2023), potentially to the extent that partial sets and mixed objects become 'good enough' referents. Secondly, our experimental design involved two key phases, an initial and a final phase, which determined the partial status of objects and sets of objects. However, a different sequence of events may influence how well partial entities fit a given description, e.g. partial objects might be perceived differently depending on whether they were intentionally created or resulted from an accident. Another promising direction for future research would be to allow children to provide more fine-grained evaluations by using a ranking procedure that includes a middle option. For instance, if participants were given three reward choices (e.g. lemon vs. tomato vs. strawberry), as in Experiment 1 of Syrett and Aravind (2022) and Experiment 2 of Tieu et al. (2019), they might have selected the medium reward in intermediate scenarios.

120

4.7 General discussion

Vagueness is a pervasive feature of natural language and extends beyond it (cf. Russell, 1923). In this study, we examined how children deal with the uncertainty arising from potentially vague or imprecise expressions. Specifically, we investigated whether their previously observed non-adult performance in interpreting definite descriptions stems from an initial expectation of sharp boundaries and clear-cut truth conditions, or whether they show sensitivity to intermediate cases.

There were two conditions in our study in which a pattern emerged that diverges from what we would expect if the respective expressions are assigned clear-cut, complementary truth-conditions. For one, we found that in the partial plural condition in Experiment 4, many children rejected both positive and negative sentences when presented with non-maximal sets of objects. Secondly, in the mixed singular condition in Experiment 5, acceptance for both positive and negative sentences was at chance level, with most children providing inconsistent judgments when presented with heterogeneous single objects.

This suggests that children when interpreting definite singular and plural descriptions were sensitive to intermediate cases which fall in between clearly true and false cases. At the same time, the pattern also suggests that our task gave rise to two different kinds of oddness, both of which were recognized by our participants. The first kind of oddness or uncertainty arose from the homogeneity property of definite plural sentences. Here many children showed an adult-like performance and treated non-maximal scenarios as neither falling under the positive nor under the negative description. The second type of uncertainty likely arose from the indeterminacy of singular count nouns. While children, unlike in the plural condition, were permissive when presented with partial objects, their symmetric judgments in the singular mixed condition suggest that heterogeneous objects were treated as borderline cases, in particular, borderline case of vague nominals (cf. Syrett and Aravind, 2022 for a similar perspective; also Kamp and Partee, 1995; Chierchia, 2010, a.o.). Thus, the fact that many children derived homogeneous interpretations of plural definites, but not of singular definites may potentially stem from their flexible

121

interpretation of singular count nouns. However, overall the asymmetric response patterns found in the non-maximal conditions of both experiments can be accounted for by implicature-based theories of homogeneity (cf. Magri, 2014; Bar-Lev, 2021) whereby the pattern suggests that many children accessed an existential meaning and some computed the implicature.

The findings are in line with previous work which suggests that young children are attuned to category variability and to specific properties of indeterminate expressions. For instance, studies have shown that by the age of 3 to 4 children recognize the context-sensitivity of vague predicates such as *big* and *tall* and adjust their interpretation accordingly (Barner and Snedeker, 2008; Syrett et al., 2010). Similarly, children adapt their interpretations of indefinite singulars based on contextual affordances, i.e. when the importance of the function of an object is made salient in the discourse, partial objects are accepted to a lesser degree than in neutral contexts (Syrett and Aravind, 2022). Regarding plural definite descriptions, Tieu et al. (2019) found that some children rejected both affirmative and negated sentences containing plural definites in non-maximal scenarios, and that at a group level these sentences were judged as less acceptable than in maximal scenarios, suggesting that children recognize the intermediate status of such scenarios.

Our study adds to these findings by demonstrating that in both the singular and the plural domain children do not always interpret definite descriptions in terms of sharp boundaries, instead they are attuned to intermediate cases. From a broader perspective, we may conclude that while categorical thinking is fundamental to human cognition (Quinn et al., 2001; Behl-Chadha, 1996; Quinn and Eimas, 1998; Pauen, 2002; Furrer and Younger, 2005, among many others) and induced by as well as reflected in language (Waxman and Markow, 1995; Balaban and Waxman, 1997; Fulkerson and Haaf, 2006, a.o.), our findings indicate that even young preschoolers can take into account the indeterminacy of certain meanings and acknowledge that categorical distinctions introduced by linguistic expressions may have blurred boundaries and may include uncertain cases.

Chapter 5

Conclusion

Linguistic competence encompasses knowledge at multiple interrelated levels of abstraction. First, knowing a language requires acquiring its lexical items along with their respective meanings. Additionally, it involves an understanding of formal principles that hold across languages, such as the structure-dependency of grammatical rules. Furthermore, linguistic competence presupposes knowledge about essential semantic and pragmatic properties of natural language, such as its symbolic nature and its role in conveying relevant meaning.

When acquiring a language, children thus face a multifaceted learnability challenge. However, arguably, not all aspects of linguistic knowledge must be learned inductively, some may be part of the innate faculty that enables language acquisition. For instance, from the outset, children expect syntactic rules to operate on hierarchical representations, as evidenced by the absence of errors suggesting reliance on linear order. Other aspects, such as the specific linguistic forms used to convey a given meaning in a particular language, will have to be learned through exposure, as languages obviously differ considerably in this regard.

This thesis addressed the question of how children come the know the meanings of plural expressions. In the first part we looked at how children acquire the distributive interpretation of ambiguous plural sentences and whether this meaning has to be learned, or constructed, along with the acquisition of distributive universal quantifying expressions. Our study shows

that children have access to a distributive universal quantifier before they have mastered the truth-conditions of the corresponding explicit linguistic form, suggesting that it is not the construction of the logical representation that causes difficulties during the acquisition process, but rather the correct mapping between form and meaning. Our findings thus provide initial evidence in favor of theories which assume that children do not need to create new representational resources when acquiring the meaning of logical expressions.

The second part was concerned with the question to what extent certain formal properties that are part of natural language meanings are also part of non-linguistic representations. Specifically, we asked whether exhaustivity is a concept available to preverbal infants and may be used to encode events that involve multiple individuals. The results from our study are inconclusive, in that many infants did clearly not make use of such a concept, while others potentially did. This may indicate that the exhaustive/non-exhaustive contrast is an early emerging conceptual distinction that is available independent of language but not frequently invoked, but it may also reflect developmental constraints on the availability of such a distinction. As such, the results do not provide decisive evidence either for or against competing theories of early conceptual development.

In the third part, we investigated whether the indeterminacy associated with linguistic expressions is something that children gradually come to acknowledge or whether there is evidence that they bring early expectations about this feature of language to the acquisition process. Our study demonstrates that at least in the domain of plurals and singulars already young preschoolers are attuned to intermediate cases, suggesting that the indeterminacy of certain linguistic expression, just like categorical distinctions, is included in children's early implicit knowledge about natural language meanings.

Taken together these findings provide evidence that children are prepared from an early age to connect certain logical meanings to their corresponding linguistic forms and to deal with the uncertainty that arises with some expressions. At the same time, they point to the possibility that the logical representations that are deployed prior to and outside of language may be built from more primitive or potentially distinct representations than those underlying language.

5.1 Implications and outlook

5.1.1 Acquiring distributive universals

In our investigation of plural ambiguities in chapter 2 we asked whether children have access to a distributive quantifier meaning before acquiring the corresponding lexical item, however, the study leaves open the question of how children come to know that this meaning is specifically linked to *each* and its equivalents in other languages, i.e. we have not addressed which types of evidence help children to narrow down their interpretation of distributive and non-distributive quantifiers. While some have proposed that conversational principles that involve competition might help children in determining the meanings of plural expressions and quantified expressions (cf. Dotlačil, 2010; Pagliarini et al., 2012), there is also evidence suggesting that children may differ in how they construe events, in particular, that they tend to construe distributive events, which involve actions performed separately by individuals, nevertheless as collective, leading them to accept collectivity markers (e.g. *together*) in such cases (Syrett and Musolino, 2016).

If children's representational capacities are in principle powerful enough to encode a distributive quantifier meaning, then one of the challenges in assigning such a meaning to distributivity markers may stem from such extra-linguistic factors, e.g. children may evaluate the evidence for a distributive interpretation of these markers differently due to their different construal of the situations in which such markers are uttered. At the same time, although already infants have been shown to be able to represent various aspects of individual and collective actions, and children as young as three have demonstrated an understanding that certain predicates apply to individuals while others hold of groups (Syrett, 2014), it remains possible that young children do not initially consider the distributive/non-distributive distinction to be linguistically relevant. Instead, this distinction may only become salient as they work to determine the meanings of distributive and non-distributive quantifiers. One cue that may help them to recognize the distinct meanings is the number marking on the nouns they combine with. Specifically,

cross-linguistic evidence suggests that distributive quantifiers typically combine with singular nouns, whereas quantifiers that also allow for non-distributive interpretations combine with plural nouns (Gil, 1995; Haslinger et al., 2023).

Besides, an interesting prediction that follows from the assumptions underlying the logical nativist perspective, which finds some support in our data, is that in principle infants and children should be able to learn any logical contrast that is expressed in natural language, regardless of them having acquired any associated linguistic material, maybe analogous to what is observed in early phonological development. A promising direction for future research would be to examine whether and how easily children acquiring a language which lacks distinct lexicalized forms for distributive and non-distributive universal quantifiers can learn to map these distinct meanings onto newly introduced forms.

5.1.2 Navigating uncertainty

Similarly, the study on definite descriptions in chapter 4 was concerned with the question whether children are sensitive to intermediate cases, and did not provide details on how children resolve uncertainty in such situations. A natural continuation of our and related work would be to explore in greater detail to what extent children (and adults) rely on contextual information, functional and intentional cues, and similarity to prototypes, and how different populations respond to varying degrees of homogeneity violations. At the same time, an important question that deserves attention is how children come to understand that certain expressions allow for imprecision in specific contexts, while others do not (or not to the same extent). Our findings indicate that children are attuned to intermediate cases when interpreting definite descriptions, leading us to speculate that the potential for indeterminacy or vagueness may be an inherent expectation about language rather than something learned through experience. On the other hand, some theories assume that definite descriptions come with precise interpretations in certain contexts, but may lead to uncertainty in others (cf. Haslinger, 2022). Understanding how children learn to distinguish such contexts, especially since this information is not encoded

in the expressions themselves, and how their preparedness for categorical as well as imprecise or vague meanings operates during this process could provide key insights into language development more broadly.

5.1.3 Early logical representations

Although the concept of exhaustivity appears to be an extremely useful if not essential component of processes such as hypothesis testing and inferential reasoning, our study on exhaustivity discussed in chapter 3 does not provide conclusive evidence regarding its availability to infants. However, one possible conclusion from this investigation is that infants are indeed capable of learning rules based on the quantificational properties of events. Yet, what remains unclear is why many of our participants did not learn any rule. In fact the non-learners in the sample consistently looked more to the rule incongruent side during both the familiarization and the test phase. We do not know how the non-learners in our sample encoded the scenes, nor whether they used a similar encoding strategy but failed to link the exhaustive/non-exhaustive feature to a particular outcome due to processing limitations. Alternatively, their gaze behavior might have been driven by exploration rather than by prediction.

This is also related to some broader limitations of such research, in particular, that we generally lack insights into the precise format of infants' representation of the scenes. Even for infants who successfully learned the rules, it remains difficult to determine the complexity of their representations and the specific elements they contain. Future studies could attempt to address these limitations by introducing a broader range of events with varying visual and conceptual properties while maintaining the logical distinction between exhaustive and non-exhaustive events. This approach would help to assess the extent to which infants generalize such patterns and to clarify whether quantification is involved. If quantification plays a role, it could be examined further what exactly infants quantify over and which distinctions are relevant (e.g. NONE vs. NOT ALL vs. ALL).

Notably, if the relevant distinction was indeed ALL vs. NOT ALL, then our findings would not

only suggest commonalities between linguistic and non-linguistic representations but also point to potential differences, since it is well known that NOT ALL is not lexicalized in any language. From this perspective, our results may contribute to the debate on linguistic universals by raising an important question, namely why a distinction that may underlie non-linguistic reasoning processes is absent from the vocabularies of the languages of the world.

Appendix A: Supplementary materials to chapter 2

A.1   Test items Experiment 1

ID	Prime	Target
1	*Zwei Gänse ziehen zwei Hasen.* 'Two geese are pulling two rabbits.'	*Zwei Schweine ziehen drei Kühe.* 'Two pigs are pulling three cows.'
2	*Zwei Gänse ziehen zwei Pferde.* 'Two geese are pulling two horses.'	*Zwei Hasen ziehen drei Kühe.* 'Two rabbits are pulling three cows.'
3	*Zwei Hasen ziehen drei Gänse.* 'Two rabbits are pulling three geese.'	*Zwei Schweine ziehen zwei Pferde.* 'Two pigs are pulling two horses.'
4	*Zwei Hasen ziehen zwei Kühe.* 'Two rabbits are pulling two cows.'	*Zwei Pferde ziehen drei Schweine.* 'Two horses are pulling three pigs.'
5	*Zwei Hasen ziehen drei Pferde.* 'Two rabbits are pulling three horses.'	*Zwei Gänse ziehen zwei Schweine.* 'Two geese are pulling two pigs.'
6	*Zwei Hasen ziehen zwei Schweine.* 'Two rabbits are pulling two pigs.'	*Zwei Kühe ziehen drei Pferde.* 'Two cows are pulling three horses.'
7	*Zwei Kühe ziehen zwei Gänse.* 'Two cows are pulling two geese.'	*Zwei Hasen ziehen drei Schweine.* 'Two rabbits are pulling three pigs.'
8	*Zwei Kühe ziehen drei Hasen.* 'Two cows are pulling three rabbits.'	*Zwei Pferde ziehen zwei Gänse.* 'Two horses are pulling two geese.'
9	*Zwei Kühe ziehen drei Schweine.* 'Two cows are pulling three pigs.'	*Zwei Hasen ziehen zwei Gänse.* 'Two rabbits are pulling two geese.'
10	*Zwei Pferde ziehen drei Gänse.* 'Two horses are pulling three geese.'	*Zwei Kühe ziehen zwei Schweine.* 'Two cows are pulling two pigs.'
11	*Zwei Pferde ziehen zwei Hasen.* 'Two horses are pulling two rabbits.'	*Zwei Kühe ziehen drei Gänse.* 'Two cows are pulling three geese.'
12	*Zwei Pferde ziehen drei Kühe.* 'Two horses are pulling three cows.'	*Zwei Schweine ziehen zwei Hasen.* 'Two pigs are pulling two rabbits.'
13	*Zwei Pferde ziehen zwei Schweine.* 'Two horses are pulling two pigs.'	*Zwei Gänse ziehen drei Hasen.* 'Two geese are pulling three rabbits.'
14	*Zwei Schweine ziehen drei Hasen.* 'Two pigs are pulling three rabbits.'	*Zwei Gänse ziehen zwei Kühe.* 'Two geese are pulling two cows.'
15	*Zwei Schweine ziehen zwei Kühe.* 'Two pigs are pulling two cows.'	*Zwei Gänse ziehen drei Pferde.* 'Two geese are pulling three horses.'
16	*Zwei Schweine ziehen drei Pferde.* 'Two pigs are pulling three horses.'	*Zwei Kühe ziehen zwei Hasen.* 'Two cows are pulling two rabbits.'

A.2   Test items Experiment 2

ID	Prime	Target
1	*Zwei Gänse schieben zwei Hasen.* 'Two geese are pushing two rabbits.'	*Zwei Schweine ziehen drei Kühe.* 'Two pigs are pulling three cows.'
2	*Zwei Gänse schieben zwei Pferde.* 'Two geese are pushing two horses.'	*Zwei Hasen ziehen drei Kühe.* 'Two rabbits are pulling three cows.'
3	*Zwei Hasen schieben drei Gänse.* 'Two rabbits are pushing three geese.'	*Zwei Schweine ziehen zwei Pferde.* 'Two pigs are pulling two horses.'
4	*Zwei Hasen schieben zwei Kühe.* 'Two rabbits are pushing two cows.'	*Zwei Pferde ziehen drei Schweine.* 'Two horses are pulling three pigs.'
5	*Zwei Hasen schieben drei Pferde.* 'Two rabbits are pushing three horses.'	*Zwei Gänse ziehen zwei Schweine.* 'Two geese are pulling two pigs.'
6	*Zwei Hasen schieben zwei Schweine.* 'Two rabbits are pushing two pigs.'	*Zwei Kühe ziehen drei Pferde.* 'Two cows are pulling three horses.'
7	*Zwei Kühe schieben zwei Gänse.* 'Two cows are pushing two geese.'	*Zwei Hasen ziehen drei Schweine.* 'Two rabbits are pulling three pigs.'
8	*Zwei Kühe schieben drei Hasen.* 'Two cows are pushing three rabbits.'	*Zwei Pferde ziehen zwei Gänse.* 'Two horses are pulling two geese.'
9	*Zwei Kühe schieben drei Schweine.* 'Two cows are pushing three pigs.'	*Zwei Hasen ziehen zwei Gänse.* 'Two rabbits are pulling two geese.'
10	*Zwei Pferde schieben drei Gänse.* 'Two horses are pushing three geese.'	*Zwei Kühe ziehen zwei Schweine.* 'Two cows are pulling two pigs.'
11	*Zwei Pferde schieben zwei Hasen.* 'Two horses are pushing two rabbits.'	*Zwei Kühe ziehen drei Gänse.* 'Two cows are pulling three geese.'
12	*Zwei Pferde schieben drei Kühe.* 'Two horses are pushing three cows.'	*Zwei Schweine ziehen zwei Hasen.* 'Two pigs are pulling two rabbits.'
13	*Zwei Pferde schieben zwei Schweine.* 'Two horses are pushing two pigs.'	*Zwei Gänse ziehen drei Hasen.* 'Two geese are pulling three rabbits.'
14	*Zwei Schweine schieben drei Hasen.* 'Two pigs are pushing three rabbits.'	*Zwei Gänse ziehen zwei Kühe.* 'Two geese are pulling two cows.'
15	*Zwei Schweine schieben zwei Kühe.* 'Two pigs are pushing two cows.'	*Zwei Gänse ziehen drei Pferde.* 'Two geese are pulling three horses.'
16	*Zwei Schweine schieben drei Pferde.* 'Two pigs are pushing three horses.'	*Zwei Kühe ziehen zwei Hasen.* 'Two cows are pulling two rabbits.'

A.3   Control items Experiment 1 & 2

ID	1st sentence	2nd sentence
17	*Da ist eine Gans.* 'There is one goose.'	*Da sind zwei Schweine.* 'There are two pigs.'
18	*Da sind zwei Hasen.* 'There are two rabbits.'	*Da ist eine Kuh.* 'There is one cow.'
19	*Da sind drei Pferde.* 'There are three horses.'	*Da ist ein Hase.* 'There is one rabbit.'
20	*Da ist ein Schwein.* 'There is one pig.'	*Da sind drei Gänse.* 'There are three geese.'
21	*Da sind zwei Kühe* 'There are two cows.'	*Da sind drei Hasen.* 'There are three rabbits.'
22	*Da sind drei Schweine.* 'There are three pigs.'	*Da sind zwei Pferde.* 'There are two horses.'

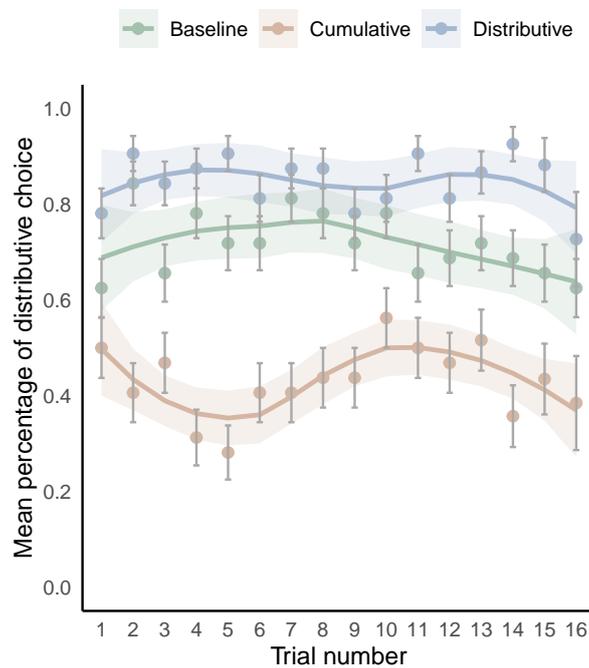## A.4 Additional figures Experiment 1 & 2

Figure A.1: **Means by trial in Exp. 1** Mean percentage of distributive choices in target trials by trial number (1 to 16) across experimental conditions in Experiment 1. Points represent the mean response per condition and trial, with error bars indicating the standard error.
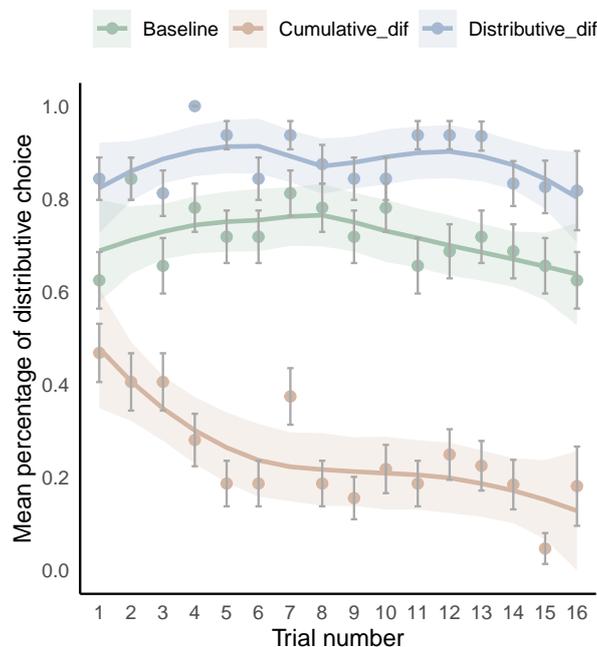

Figure A.2: **Means by trial in Exp. 2** Mean percentage of distributive choices in target trials by trial number (1 to 16) across experimental conditions in Experiment 2. Points depict the mean response per condition and trial, with error bars indicating the standard error.

Appendix B: Supplementary materials to chapter 4

B.1 Test items Experiment 4

ID	Positive sentences	Negative sentences
1	*Jack hat die Hämmer gefunden.* 'Jack found the hammers.'	*Jack hat die Hämmer nicht gefunden.* 'Jack didn't find the hammers.'
2	*Jack hat die Messer gefunden.* 'Jack found the knives.'	*Jack hat die Messer nicht gefunden.* 'Jack didn't find the knives.'
3	*Jack hat die Brillen gefunden.* 'Jack found the (pairs of) glasses.'	*Jack hat die Brillen nicht gefunden.* 'Jack didn't find the (pairs of) glasses.'
4	*Jack hat die Scheren gefunden.* 'Jack found the (pairs of) scissors.'	*Jack hat die Scheren nicht gefunden.* 'Jack didn't find the (pairs of) scissors.'
5	*Jack hat die Gläser gefunden.* 'Jack found the glasses.'	*Jack hat die Gläser nicht gefunden.* 'Jack didn't find the glasses.'
6	*Jack hat die Gabeln gefunden.* 'Jack found the forks.'	*Jack hat die Gabeln nicht gefunden.* 'Jack didn't find the forks.'
7	*Jack hat die Schneebesen gefunden.* 'Jack found the whisks.'	*Jack hat die Schneebesen nicht gefunden.* 'Jack didn't find the whisks.'
8	*Jack hat die Pinsel gefunden.* 'Jack found the the brushes.'	*Jack hat die Pinsel nicht gefunden.* 'Jack didn't find the brushes.'
9	*Jack hat die Gitarren gefunden.* 'Jack found the guitars.'	*Jack hat die Gitarren nicht gefunden.* 'Jack didn't find the guitars.'
10	*Jack hat die Staubsauger gefunden.* 'Jack found the vacuum cleaners.'	*Jack hat die Staubsauger nicht gefunden.* 'Jack didn't find the vacuum cleaners.'
11	*Jack hat die Trompeten gefunden.* 'Jack found the trumpets.'	*Jack hat die Trompeten nicht gefunden.* 'Jack didn't find the trumpets.'
12	*Jack hat die Besen gefunden.* 'Jack found the brooms.'	*Jack hat die Besen nicht gefunden.* 'Jack didn't find the brooms.'
13	*Jack hat die Roller gefunden.* 'Jack found the scooters.'	*Jack hat die Roller nicht gefunden.* 'Jack didn't find the scooters.'
14	*Jack hat die Lampen gefunden.* 'Jack found the lamps.'	*Jack hat die Lampen nicht gefunden.* 'Jack didn't find the lamps.'
15	*Jack hat die Kleiderbügel gefunden.* 'Jack found the hangers.'	*Jack hat die Kleiderbügel nicht gefunden.* 'Jack didn't find the hangers.'
16	*Jack hat die Zahnbürsten gefunden.* 'Jack found the toothbrushes.'	*Jack hat die Zahnbürsten nicht gefunden.* 'Jack didn't find the toothbrushes.'

B.2   Control items Experiment 4

ID	Control sentences

17   *Jack hat die Uhren gefunden.*
'Jack found the clocks.'

18   *Jack hat die Teekannen gefunden.*
'Jack found the teapots.'

19   *Jack hat die Piratenhüte nicht gefunden.*
'Jack didn't find the pirate hats.'

20   *Jack hat die Töpfe nicht gefunden.*
'Jack didn't find the pots.'

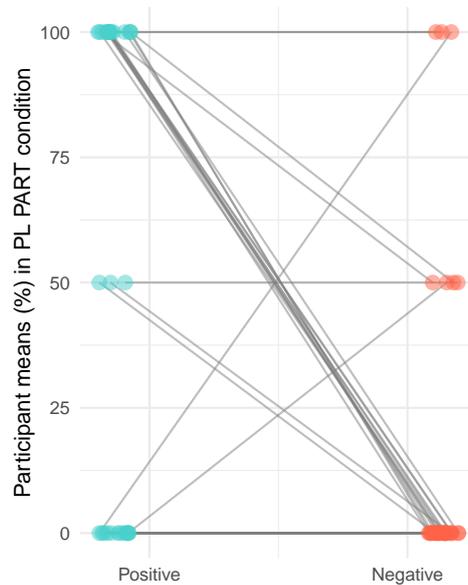# B.3 Additional figures Experiment 4

Figure B.1: **Participant means in partial condition in Exp. 4** Mean percentage of yes-responses per participant in the partial condition in Experiment 4. Each dot represents an individual participant's mean in the respective polarity, with lines connecting each participant's means across the two polarity condition.
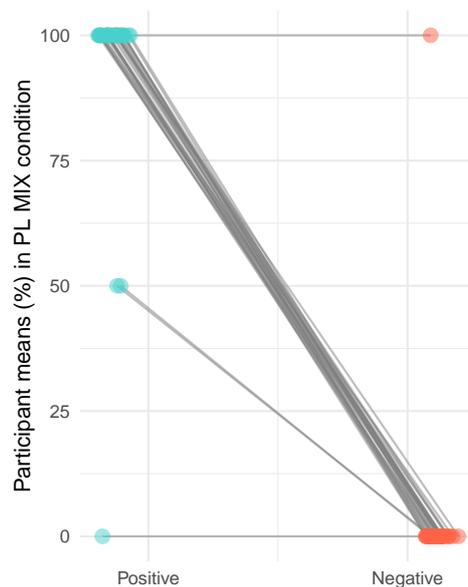

Figure B.2: **Participant means in mixed condition in Exp. 4** Mean percentage of yes-responses per participant in the mixed condition in Experiment 4. Dots represent individual participant's means and lines connect individual means across the two polarity conditions.

B.4 Test items Experiment 5

ID	Positive sentences	Negative sentences
1	*Jack hat den Hammer gefunden.* 'Jack found the hammer.'	*Jack hat den Hammer nicht gefunden.* 'Jack didn't find the hammer.'
2	*Jack hat das Messer gefunden.* 'Jack found the knife.'	*Jack hat das Messer nicht gefunden.* 'Jack didn't find the knife.'
3	*Jack hat die Brille gefunden.* 'Jack found the (pair of) glasses.'	*Jack hat die Brille nicht gefunden.* 'Jack didn't find the (pair of) glasses.'
4	*Jack hat die Schere gefunden.* 'Jack found the (pair of) scissors.'	*Jack hat die Schere nicht gefunden.* 'Jack didn't find the (pair of) scissors.'
5	*Jack hat das Glas gefunden.* 'Jack found the glass.'	*Jack hat das Glas nicht gefunden.* 'Jack didn't find the glass.'
6	*Jack hat die Gabel gefunden.* 'Jack found the fork.'	*Jack hat die Gabel nicht gefunden.* 'Jack didn't find the fork.'
7	*Jack hat den Schneebesen gefunden.* 'Jack found the whisk.'	*Jack hat den Schneebesen nicht gefunden.* 'Jack didn't find the whisk.'
8	*Jack hat den Pinsel gefunden.* 'Jack found the the brush.'	*Jack hat den Pinsel nicht gefunden.* 'Jack didn't find the brush.'
9	*Jack hat die Gitarre gefunden.* 'Jack found the guitar.'	*Jack hat die Gitarre nicht gefunden.* 'Jack didn't find the guitar.'
10	*Jack hat den Staubsauger gefunden.* 'Jack found the vacuum cleaner.'	*Jack hat den Staubsauger nicht gefunden.* 'Jack didn't find the vacuum cleaner.'
11	*Jack hat die Trompete gefunden.* 'Jack found the trumpet.'	*Jack hat die Trompete nicht gefunden.* 'Jack didn't find the trumpet.'
12	*Jack hat den Besen gefunden.* 'Jack found the broom.'	*Jack hat den Besen nicht gefunden.* 'Jack didn't find the broom.'
13	*Jack hat den Roller gefunden.* 'Jack found the scooter.'	*Jack hat den Roller nicht gefunden.* 'Jack didn't find the scooter.'
14	*Jack hat die Lampe gefunden.* 'Jack found the lamp.'	*Jack hat die Lampe nicht gefunden.* 'Jack didn't find the lamp.'
15	*Jack hat den Kleiderbügel gefunden.* 'Jack found the hanger.'	*Jack hat den Kleiderbügel nicht gefunden.* 'Jack didn't find the hanger.'
16	*Jack hat die Zahnbürste gefunden.* 'Jack found the toothbrush.'	*Jack hat die Zahnbürste nicht gefunden.* 'Jack didn't find the toothbrush.'

B.5   Control items Experiment 5

ID	Control sentences

17   *Jack hat die Uhr gefunden.*
'Jack found the clock.'

18   *Jack hat die Teekanne gefunden.*
'Jack found the teapot.'

19   *Jack hat den Piratenhut nicht gefunden.*
'Jack didn't find the pirate hat.'

20   *Jack hat den Topf nicht gefunden.*
'Jack didn't find the pot.'

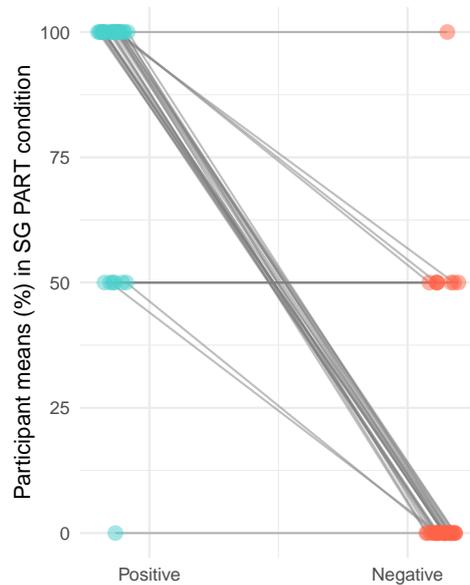# B.6   Additional figures Experiment 5

Figure B.3: **Participant means in partial condition in Exp. 5** Mean percentage of yes-responses per participant in the partial condition in Experiment 5. Each dot represents an individual participant's mean in the respective polarity and lines connect individual means across the two polarity condition.
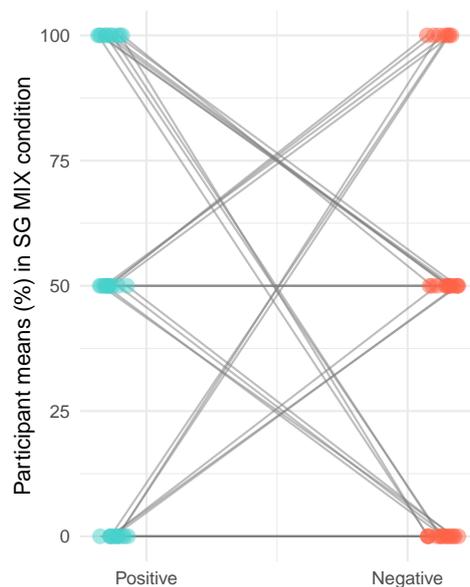

Figure B.4: **Participant means in mixed condition in Exp. 5** Mean percentage of yes-responses per participant in the mixed condition in Experiment 5. Dots represent individual participant's means and lines connect individual means across the polarity conditions.

138

Appendix C: Conjunctive plural sentences

In this study we investigated conjunctive plural sentences such as (69). Plural sentences that contain a conjunction of two (or more) predicates have been observed to allow for a distributive construal, on which both properties apply to each member of the subject plurality as in (70a), but also for a cumulative construal, where some members of the subject plurality have one, and some the other property as in (70b) (cf. Link, 1983; Krifka, 1990; Schmitt, 2013; Winter, 2001; Poortman, 2014, a.o.).

(69) *The six girls are sitting and cooking.*

(70) a. DISTRIBUTIVE SCENARIO: The six girls are sitting and the six girls are cooking.

   b. CUMULATIVE SCENARIO: Three of the girls are sitting, the other three are cooking.

While some theories assume that the basic meaning of such sentences is cumulative and predict both construals to be available (Link, 1983; Krifka, 1990; Schmitt, 2013), alternative accounts propose that the default interpretation is distributive and that cumulative readings arise only when supported by world-knowledge. In particular, cumulative construals become available when the conjuncts denote disjoint properties, i.e. properties which cannot hold simultaneously of a single individual (e.g. sitting and standing) (Winter, 2001), or when their co-occurrence in a single individual is atypical (e.g. sitting and cooking) (Poortman, 2014). We examined conjunctive plural sentences in Hungarian such as (71), which are additionally ambiguous between a specific and a generic reading.

(71) *A   malabik csíkosak és   szárnyaik   vannak.*
    DET malabi.PL striped.PL AND wing.POSS.3PL be.3PL

   'The malabis are striped and have wings.' (specific)
   'Malabis are striped and have wings.' (generic)

By constructing a context that favored the generic interpretation, specifically, by embedding the test sentences within a story about newly discovered animal kinds, we aimed to investigate whether speakers would access the cumulative construal more readily even for properties which are not disjoint. We hypothesized that in such cases world-knowledge, particularly knowledge about the distribution of properties within kinds (e.g. female and male members of a kind), might support a cumulative interpretation of properties that can in principle simultaneously hold of a single individual. In Experiment 6, we measured the acceptability of plural sentences such as (71) in non-maximal distributive and cumulative scenarios, and compared these to 'most'-quantified sentences such as (72), which were expected to be acceptable only in distributive scenarios.

(72) *A   legtöbb malabi   csíkos   és   szárnyai   vannak.*
    DET most   malabi.SG striped.SG AND wing.POSS.3SG be.3PL

   'Most malabis are striped and have wings.'

In Experiment 7, the acceptability of plural sentences in maximal distributive and cumulative scenarios was assessed, and compared to 'every'-quantified sentences such as (73), which again

were not expected to permit cumulative construals.

(73) *Minden malabi csíkos és szárnyai vannak.*
every malabi.SG striped.SG AND wing.POSS.3SG be.3PL

'Every malabi is striped and has wings.'

Overall the study showed that the acceptability of conjunctive plural sentences was not affected by the non-maximal/maximal distinction, however, it did not reveal a clear difference between conjunctive plural and quantified sentences with respect to the cumulative/distributive contrast. Although the acceptability of plural sentences was slightly increased in cumulative scenarios compared to quantified sentences, the response rates did not differ from chance, suggesting that participants did not reliably access cumulative interpretations.

C.1 Experiment 6: Plural vs. MOST

C.1.1 Methods

Participants

We tested 32 native Hungarian-speaking children (range = 5;01 – 5;11 years, *M* = 5;06 years, *SD* = 0.24, 18 females). 2 additional children were tested but excluded from the final sample because they did not meet our inclusion criteria. Participants were tested via Zoom in the presence of a caretaker. The study was approved by the United Ethical Review Committee for Research in Psychology in Hungary (EPKEB) and informed consent was obtained before the testing session.

Design

The study involved 2 (subject: plural, MOST) x 2 (scenario: distributive, cumulative) conditions within-subjects and used a Truth-Value Judgment Task (TVJT). All sentences were embedded within a coherent story about a space traveler named Laura, who recorded her observations in a diary. In each trial, participants were presented with a sentence from Laura's diary and asked to judge whether it accurately described an accompanying picture. In the plural condition, sentences contained a definite plural subject combined with two conjoined predicates (e.g. 'The malabis/Malabis are striped and have wings'). In the MOST condition, the subject phrase was quantified by 'most' and combined with two conjoined predicates (e.g. 'Most malabis are striped and have wings'). In distributive scenarios, pictures depicted 8 creatures of which 6 exhibited both properties mentioned in the sentence (e.g. being striped and having wings), while 2 exhibited neither. In cumulative scenarios, 8 creatures were shown of which 3 exhibited one of the properties (e.g. being striped), another 3 exhibited the other property (e.g. having wings), and the remaining 2 had neither of these properties. Thus, all test scenarios in Experiment 6 were non-maximal as each included two creatures that exhibited neither of the properties mentioned in the sentence. The dependent measure was the rate of yes-responses

140

in test trials. Familiarization trials followed the same structure, but involved definite singular sentences containing one predicate (e.g. 'The planet is blue') shown together with a picture depicting a single object or character.
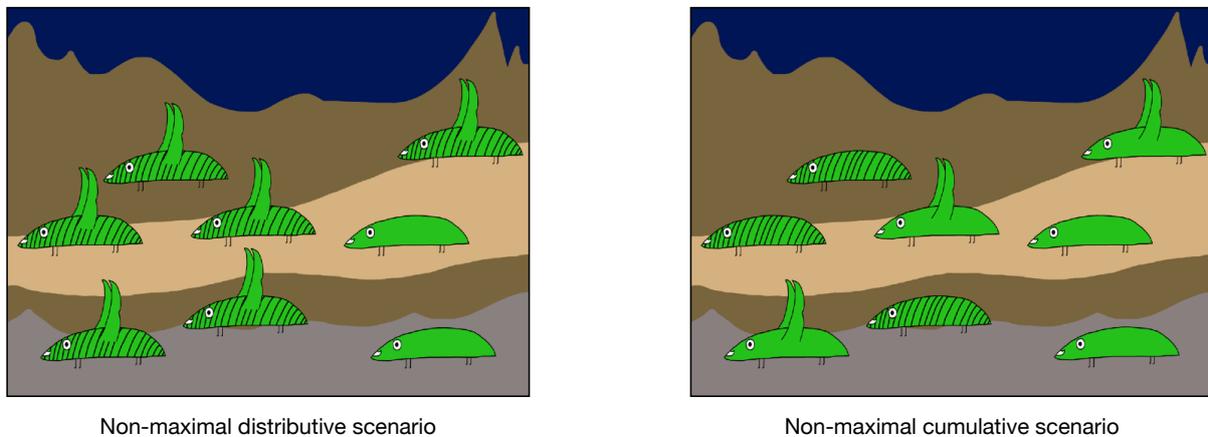

Non-maximal distributive scenario   Non-maximal cumulative scenario

Figure C.1: Example stimuli in Exp. 6 for the predicates 'being striped and having wings'

Materials

We used 8 pseudowords denoting unfamiliar animal categories to construct our test sentences. Each noun was paired with one of 8 distinct combinations of properties, resulting in a total of 8 test sentences. Test sentences contained either a plural or a 'most'-quantified phrase in subject position and followed one of the two structures below.

(74) a.   The *X*s are *P* and *Q*.

  b.   Most *X*s are *P* and *Q*.

An additional set of 8 sentences was constructed for the familiarization phase. These sentences included different noun phrase and predicates, each containing a definite singular and a single predicate, and conformed to the following structure:

(75) The *X* is *P*.

For the test phase 16 pictures were created depicting either a non-maximal distributive or cumulative scenario. Additional 8 pictures, each showing a single object or character, were used during the familiarization phase. Each participant completed 8 familiarization trials followed by 8 test trials, consisting of 2 trials per condition: distributive plural, cumulative plural, distributive MOST and cumulative MOST, presented in pseudo-random order. The combination of subject type and scenario type was counterbalanced across participants, resulting in 8 unique lists.

Procedure

Testing took place via Zoom and participants' responses were recorded and later double-coded. At the beginning of the experiment children were introduced to the main character, Laura the

space traveler. They were told that Laura had discovered many strange and funny creatures during her space travels and had taken notes in her diary to describe them. However, because she was sometimes far away, she was not always sure if her descriptions were accurate. The experimenter then asked the children if they would be willing to look at pictures of the creatures and help to decide whether Laura provided accurate descriptions of those. Following 8 familiarization trials, participants completed 8 test trials presented in pseudo-random order. In each trial, the experimenter read a sentence aloud while a corresponding picture appeared on the screen. At the end of each trial, children were asked whether they agreed with Laura's description.

C.1.2 Results

Children who did not reach an accuracy rate of at least 75% (i.e. 6 out of 8) in familiarization trials were excluded ($n = 1$). Furthermore, we excluded participants who gave only positive or only negative responses in test trials ($n = 1$). The accuracy rate in familiarization trials reached 94.14% (SE = 1.37). At test, plural sentences were accepted in distributive scenarios in 90.63% (SE = 4.16) of the cases, while the acceptance rate for plural sentences in cumulative scenarios was 40.63% (SE = 7.25). Sentences with MOST were accepted in distributive scenarios in 93.75% (SE = 3.72) of the trials, whereas the acceptance rate for MOST sentences in cumulative scenarios was 20.31% (SE = 5.43).

The data was further analyzed with logistic mixed-effects models (Jaeger, 2008) in R (v4.0.5, R Core Team, 2021) using the lme4 package (Bates et al., 2015). We fit a model with Subject (2 levels: plural, MOST), Scenario (2 levels: distributive, cumulative) and their interaction as fixed effects and included random intercepts for Participants and Items. To analyze the significance of the fixed effects a model comparisons was conducted whereby the respective effect was removed and χ^2 likelihood ratio tests were used to obtain p-values.

We compared the full model to a simpler one without Subject as predictor to test whether the rate of yes-responses was affected by Subject type. The analysis revealed a main effect of Subject ($\chi^2(1) = 4.10$, $p = .043$). The effect of Scenario was also significant ($\chi^2(1) = 132.16$, $p < .001$). There was additionally a significant interaction between Subject and Scenario ($\chi^2(1) = 5.07$, $p = .024$), indicating that in cumulative scenarios, plural sentences were significantly more likely to receive a yes-response than MOST sentences (OR = 4.15, 95% CI [1.53, 11.29], $p = .006$).

To test whether response rates in each condition differed from chance (50%), we fit a series of generalized linear mixed-effects models with that Condition (4 levels: distributive plural, cumulative plural, distributive MOST, cumulative MOST) as the reference level. In each model, the intercept represented the log-odds of a yes-response in that condition. Participants were significantly less likely than chance to provide a yes-response in the cumulative MOST condition (OR = 0.13, 95% CI [0.04, 0.43], $p = .001$). In contrast, participants were significantly more likely than chance to give a positive response in the distributive MOST condition (OR = 46.93, 95% CI [8.79, 250.68], $p < .001$) and in the distributive plural condition (OR = 25.90, 95% CI [6.16, 108.90], $p < .001$). The cumulative plural condition did not significantly differ from chance (OR = 0.53, 95% CI [0.18, 1.55], $p = .245$).
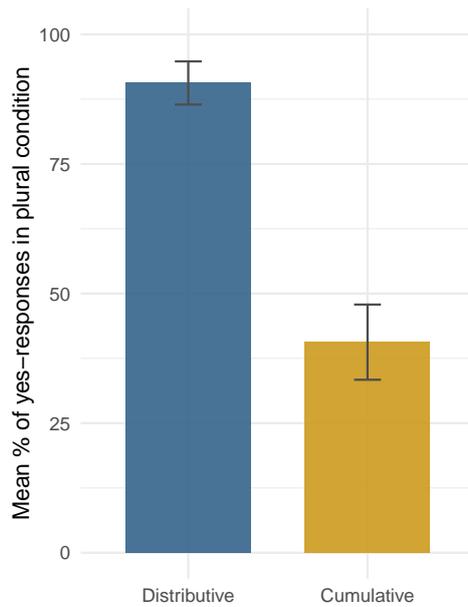
Figure C.2: **Means in plural condition in Exp. 6** Mean percentage of yes-responses in plural trials provided by participants in non-maximal distributive and cumulative scenarios in Experiment 6, with error bars depicting the standard error.
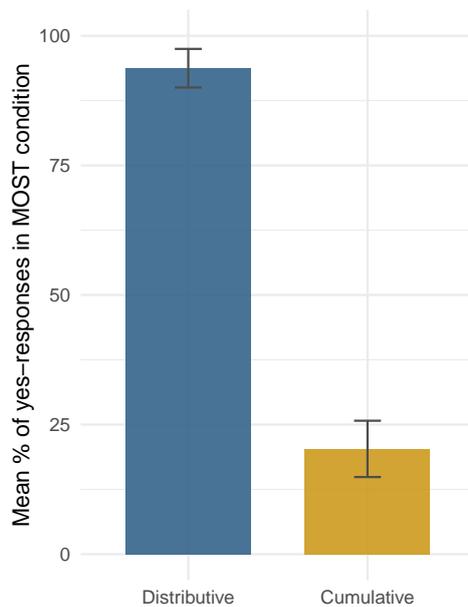

Figure C.3: **Means in MOST condition in Exp. 6** Mean percentage of yes-responses in MOST trials made by participants in non-maximal distributive and cumulative scenarios in Experiment 6. Error bars represent the standard error.

C.2 Experiment 7: Plural vs. EVERY

C.2.1 Methods

Participants

32 native Hungarian-speaking children (range = 5;00 – 5;10 years, *M* = 5;06 years, *SD* = 0.24, 18 females) participated in Experiment 7. Another 16 children were tested but had to be excluded due to not meeting our inclusion criteria. Participants were tested via Zoom, with a caretaker present. The study was approved by the United Ethical Review Committee for Research in Psychology in Hungary (EPKEB) and informed consent was obtained prior to the testing session.

Design

The experiment followed a similar design as Experiment 6 and involved 2 (subject: plural, EVERY) x 2 (scenario: distributive, cumulative) conditions within-subjects and a Truth-Value Judgment Task (TVJT). This time test sentences contained either a plural or a universally quantified noun phrase in subject position and two conjoined predicates (e.g. 'Every malabi is striped and has wings'), and test pictures displayed maximal distributive or maximal cumulative scenarios. In distributive scenarios, pictures depicted 8 creatures all of which had both properties mentioned in the sentence (e.g. being striped and having wings), while cumulative scenarios featured 8 creatures of which 4 exhibited one of the properties (e.g. being striped) and 4 exhibited the other property (e.g. having wings). The dependent measure was again the proportion of yes-responses in test trials. We included the same familiarization trials as in Experiment 6.

Materials

We used the same 8 pseudowords from Experiment 6 to construct our test sentences. Each noun was paired with a unique combinations of two properties, yielding a total of 8 test sentences. Test sentences contained either a plural or a 'every'-quantified phrase in subject position and followed one of these structures:

(76) a. The *X*s are *P* and *Q*.

 b. Every *X* is *P* and *Q*.

In addition, we created 16 test pictures depicting either a maximal distributive or a maximal cumulative scenario. As in Experiment 6, each participant completed 8 familiarization trials followed by 8 test trials (2 distributive plural, 2 cumulative plural, 2 distributive EVERY, 2 cumulative EVERY) presented in pseudo-random order.
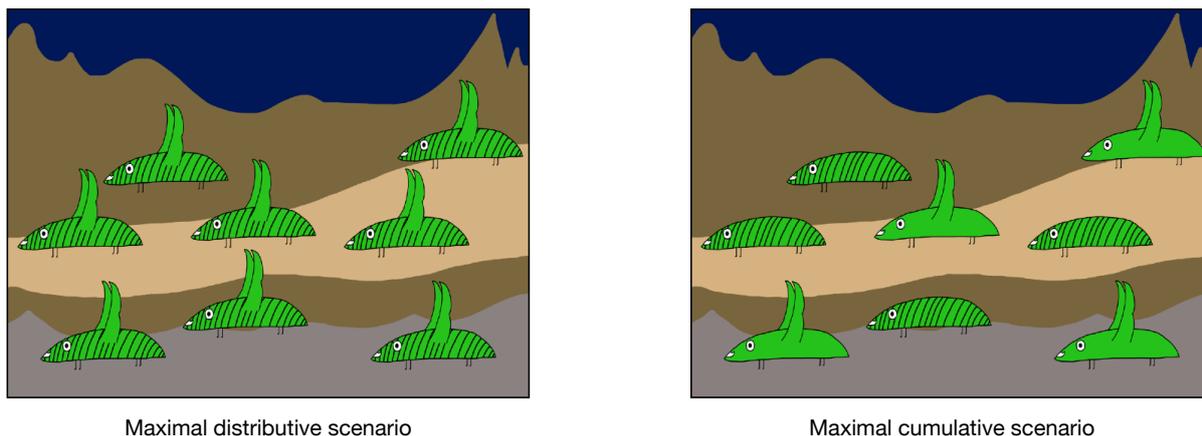
Maximal distributive scenario   Maximal cumulative scenario

Figure C.4: Example stimuli in Exp. 7 for the predicates 'being striped and having wings'

Procedure

The procedure was the same as in Experiment 6.

C.2.2 Results

Our inclusion criterion was an accuracy rate of at least 75% (i.e. 6 out of 8) in familiarization trials. All children in our sample met this criterion. We further excluded participants who gave only positive or only negative responses in test trials ($n = 16$) from the analysis. The accuracy rate in the remaining sample in familiarization trials was again high with 93.36% (SE = 1.68). The acceptance rate for plural sentences in distributive scenarios was 90.63% (SE = 4.16). In cumulative scenarios plural sentences were accepted in 42.19% (SE = 6.78) of the cases. Sentences containing EVERY were accepted in distributive scenarios in 93.75% (SE = 3.72) of trials, while the acceptance rate for EVERY sentences in cumulative scenarios was 25.0% (SE = 6.74).

We used the same statistical methods as in Experiment 6 and fit a model with Subject (2 levels: plural, EVERY), Scenario (2 levels: distributive, cumulative) and their interaction as fixed effects and included random intercepts for Participants and Items. To analyze the significance of the fixed effects a model comparisons was conducted whereby the respective effect was removed and χ^2 likelihood ratio tests were used to receive p-values.

The analysis revealed that the effect of Subject was not significant ($\chi^2(1) = 2.78$, $p = .096$), while the effect of Scenario reached significance ($\chi^2(1) = 117.38$, $p < .001$). There was no interaction between Subject and Scenario ($\chi^2(1) = 3.50$, $p = .062$).

Participants were significantly less likely than chance to give a yes-response in the cumulative EVERY condition (OR = 0.22, 95% CI $[0.09, 0.54]$, $p = .001$). In contrast, participants were significantly more likely than chance to provide a positive response in the distributive EVERY condition (OR = 31.48, 95% CI $[8.16, 121.42]$, $p < .001$) and in the distributive plural condition (OR = 18.73, 95% CI $[5.82, 60.33]$, $p < .001$). Responses in the cumulative plural condition
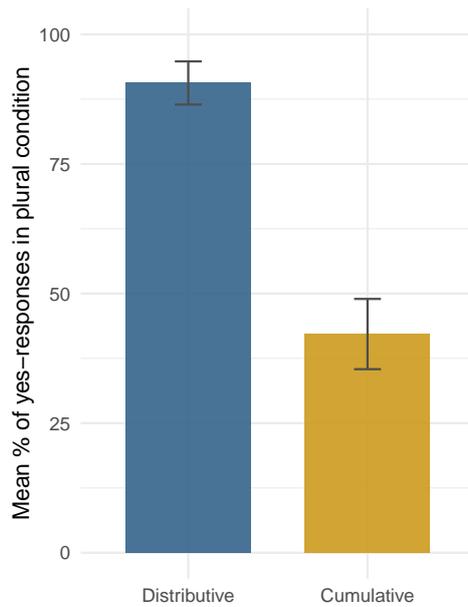
Figure C.5: **Means in plural condition in Exp. 7** Mean percentage of yes-responses in plural trials provided by participants in maximal distributive and cumulative scenarios in Experiment 7, with error bars representing the standard error.

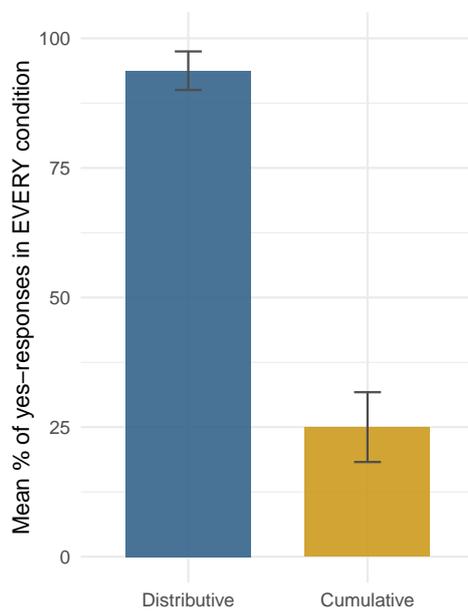did not differ significantly from chance (OR = 0.64, 95% CI $[0.29, 1.40]$, $p = .267$).

Figure C.6: **Means in EVERY condition in Exp. 7** Mean percentage of yes-responses in EVERY trials made by participants in maximal distributive and cumulative scenarios in Experiment 7. Error bars indicate the standard error.

Bibliography

Addyman, C. and D. Mareschal (2010). The perceptual origins of the abstract same/different concept in human infants. *Animal Cognition 13*(6), 817–833.

Albareda-Castellot, B., F. Pons, and N. Sebastián-Gallés (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science 14*(2), 395–401.

Altreuter, E. and J. de Villiers (2006). Does every child produce "every" correctly? In T. Heizmann (Ed.), *Current Issues in First Language Acquisition*, pp. 1–20. Amherst, MA: Graduate Linguistics Student Association (GLSA).

Alxatib, S. and F. J. Pelletier (2011). The psychology of vagueness: Borderline cases and contradictions. *Mind & Language 26*(3), 287–326.

Aravind, A., J. de Villiers, P. de Villiers, C. J. Lonigan, B. M. Phillips, J. Clancy, S. H. Landry, P. R. Swank, M. Assel, H. B. Taylor, N. Eisenberg, T. Spinrad, and C. Valiente (2017). Children's quantification with *every* over time. *Glossa: A journal of general linguistics 2*(1), 1–16.

Aravind, A. and K. Syrett (2017). *Investigating context sensitivity and vagueness in nominals in child and adult language*. Poster presented at the 48th Annual Meeting of the North East Linguistics Society (NELS). University of Iceland, Reykjavik, Iceland, October 27-29.

Augurzky, P., M. Bonnet, R. Breheny, A. Cremers, C. Ebert, C. Mayr, J. Romoli, M. Steinbach, and Y. Sudo (2023). Putting plural definites into context. In M. Onoeva, A. Staňková, and R. Šimík (Eds.), *Proceedings of Sinn und Bedeutung 27*, pp. 19–32.

Autodesk (2020). *Maya* (3D computer graphics application). https://autodesk.com/maya.

Balaban, M. T. and S. R. Waxman (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology 64*(1), 3–26.

Bar-Lev, M. E. (2021). An implicature account of homogeneity and non-maximality. *Linguistics and Philosophy 44*(5), 1045–1097.

Barner, D. and J. Snedeker (2008). Compositionality and statistics in adjective acquisition: 4-year-olds interpret *tall* and *short* based on the size distributions of novel noun referents. *Child Development 79*(3), 594–608.

Barner, D., D. Thalwitz, J. Wood, S.-J. Yang, and S. Carey (2007). On the relation between the acquisition of singular–plural morpho-syntax and the conceptual distinction between one and more than one. *Developmental Science 10*(3), 365–373.

Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language 68*(3), 255–278.

Barwise, J. and R. Cooper (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy 4*, 159–219.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1), 1–48.

Beck, S. and U. Sauerland (2000). Cumulation is needed: A reply to Winter (2000). *Natural Language Semantics 8*(4), 349–371.

Begus, K., A. Curioni, G. Knoblich, and G. Gergely (2020). Infants understand collaboration: Neural evidence for 9-month-olds' attribution of shared goals to coordinated joint actions. *Social Neuroscience 15*(6), 655–667.

Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition 60*(2), 105–141.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology 18*(3), 355–387.

Bock, K. and Z. M. Griffin (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General 129*(2), 177.

Bohus, K. A., N. Cesana-Arlotti, A. Martín-Salguero, and L. L. Bonatti (2023). The scope and role of deduction in infant cognition. *Current Biology 33*(18), 4014–4020.

Bonawitz, E., P. Shafto, Y. Yu, A. Gonzalez, and S. Bridgers (2020). Children change their answers in response to neutral follow-up questions by a knowledgeable asker. *Cognitive Science 44*(1), e12811.

Bonini, N., D. Osherson, R. Viale, and T. Williamson (1999). On the psychology of vague predicates. *Mind & Language 14*(4), 377–393.

Bott, L. and E. Chemla (2016). Shared and distinct mechanisms in deriving linguistic enrichment. *Journal of Memory and Language 91*, 117–140.

Braine, M. D. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review 85*(1), 1–21.

Branigan, H. P. and M. J. Pickering (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences 40*, e282.

Brisson, C. (1998). *Distributivity, Maximality, and Floating Quantifiers*. Ph. D. thesis, Rutgers University, New Brunswick; Piscataway.

Brisson, C. (2003). Plurals, *all*, and the nonuniformity of collective predication. *Linguistics and Philosophy 26*, 129–184.

Brooks, N., A. Pogue, and D. Barner (2011). Piecing together numerical language: Children's use of default units in early counting and quantification. *Developmental Science 14*(1), 44–57.

Brooks, P. J., M. D. Braine, G. Jia, and M. da Graca Dias (2001). Early representations for *all*, *each*, and their counterparts in Mandarin Chinese and Portuguese. In M. Bowerman and S. C. Levinson (Eds.), *Language Acquisition and Conceptual Development*, pp. 316 – 339. Cambridge: Cambridge University Press.

Brooks, P. J. and M. D. S. Braine (1996). What do children know about the universal quantifiers *all* and *each*? *Cognition 60*(3), 235–268.

Brooks, P. J. and I. Sekerina (2006). Shortcuts to quantifier interpretation in children and adults. *Language Acquisition 13*(3), 177–206.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Burnett, H. (2017). *Gradability in Natural Language. Logical and Grammatical Foundations*. Oxford: Oxford University Press.

Burnett, H. and P. R. Sutton (2020). Vagueness and Natural Language Semantics. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, and T. Zimmermann (Eds.), *The Wiley Blackwell Companion to Semantics*, pp. 1–34. Oxford: John Wiley & Sons, Ltd.

Canudas-Grabolosa, I., A. Martín-Salguero, and L. L. Bonatti (2023). Natural logic and baby LoTH. Comment on "The best game in town: The reemergence of the language-of-thought hypothesis across the cognitive sciences" by J. Quilty-Dunn, N. Porot and E. Mandelbaum. *Behavioral and Brain Sciences 46*, e266.

Caponigro, I., L. Pearl, N. Brooks, and D. Barner (2012). Acquiring the meaning of free relative clauses and plural definite descriptions. *Journal of Semantics 29*(2), 261–293.

Carey, S. (2009). *The Origin of Concepts*. New York: Oxford University Press.

Carey, S. (2011). Précis of the Origin of Concepts. *Behavioral and Brain Sciences 34*(3), 113–124.

Cesana-Arlotti, N., T. Knowlton, J. Lidz, and J. Halberda (2020). *Who's chasing whom? Adults' and infants' engagement of quantificational concepts ("Each" and "All") when representing visual chasing events*. Poster presented at the Virtual Vision Science Society Meeting, June 19–24.

Cesana-Arlotti, N., Á. M. Kovács, and E. Téglás (2020). Infants recruit logic to learn about the social world. *Nature Communications 11*(1), 5999.

Cesana-Arlotti, N., A. Martín, E. Téglás, L. Vorobyova, R. Cetnarski, and L. L. Bonatti (2018). Precursors of logical reasoning in preverbal human infants. *Science 359*(6381), 1263–1266.

Champollion, L. (2020). Distributivity, Collectivity, and Cumulativity. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, and T. Zimmermann (Eds.), *The Wiley Blackwell Companion to Semantics*, pp. 1–38. Oxford: John Wiley & Sons, Ltd.

Chierchia, G. (2010). Mass nouns, vagueness and semantic variation. *Synthese 174*, 99–149.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Clark, E. V. and T. V. Nikitina (2009). One vs. more than one: Antecedents to plural marking in early language acquisition. *Linguistics 47*(1), 103–139.

Cleland, A. A. and M. J. Pickering (2006). Do writing and speaking employ the same syntactic representations? *Journal of Memory and Language 54*(2), 185–198.

Corbett, G. G. (2000). *Number*. Cambridge: Cambridge University Press.

Crain, S. (2012). *The Emergence of Meaning*. Cambridge: Cambridge University Press.

Crain, S. and R. Thornton (1998). *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge, MA: MIT Press.

Crain, S., R. Thornton, C. Boster, L. Conway, D. Lillo-Martin, and E. Woodams (1996). Quantification without qualification. *Language Acquisition 5*(2), 83–153.

Cremers, A., M. Križ, and E. Chemla (2017). Probability judgments of gappy sentences. In S. Pistoia-Reda and F. Domaneschi (Eds.), *Linguistic and Psycholinguistic Approaches on Implicatures and Presuppositions*, Palgrave Studies in Pragmatics, Language and Cognition, pp. 111–150. London: Palgrave.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition 107*(2), 705–717.

Davies, B., N. Xu Rattanasone, and K. Demuth (2017). Two-year-olds' sensitivity to inflectional plural morphology: Allomorphic effects. *Language Learning and Development 13*(1), 38–53.

Dotlačil, J. (2010). *Anaphora and Distributivity: A study of same, different, reciprocals and others*. Ph. D. thesis, Utrecht University.

Drozd, K. (1996). Quantifier interpretation errors as errors of distributive scope. In A. Stringfellow, D. Cahana-Amitay, E. Hughes, and A. Zukowski (Eds.), *Proceedings of the 20th Annual Boston University Conference on Language Development (BUCLD)*, pp. 177–188. Somerville, MA: Cascadilla Press.

Drozd, K., D. Anđelković, M. Savić, O. Tošković, A. Gavarró, A. Lite, G. Hržica, M. Kovačević, J. Kuvač Kraljević, A. Skordi, et al. (2019). A crosslinguistic study of symmetrical judgments. In M. M. Brown and B. Dailey (Eds.), *Proceedings of the 43rd Annual Boston University Conference on Language Development (BUCLD)*, pp. 217–230. Somerville, MA: Cascadilla Press.

Drozd, K. F. (2001). Children's weak interpretations of universally quantified questions. In M. Bowerman and S. Levinson (Eds.), *Language Acquisition and Conceptual Development*, pp. 340–376. Cambridge: Cambridge University Press.

Drozd, K. F. and E. v. Loosbroek (2006). The effect of context on children's interpretations of universally quantified sentences. In V. van Geenhoven (Ed.), *Semantics in Acquisition*, Volume 35, pp. 115–140. Dordrecht: Springer.

É. Kiss, K. and T. Zétényi (2017). Quantifier spreading: children misled by ostensive cues. *Glossa: A journal of general linguistics 2*(1), 1–20.

Ekramnia, M., J. Mehler, and G. Dehaene-Lambertz (2021). Disjunctive inference in preverbal infants. *Iscience 24*(10), 103203.

Fawcett, C. and G. Gredebäck (2013). Infants use social context to bind actions into a collaborative sequence. *Developmental Science 16*(6), 841–849.

Fawcett, C. and G. Gredebäck (2015). Eighteen-month-olds, but not 14-month-olds, use social context to bind action sequences. *Infancy 20*(1), 115–125.

Feigenson, L. and S. Carey (2003). Tracking individuals via object-files: evidence from infants' manual search. *Developmental Science 6*(5), 568–584.

Feigenson, L. and S. Carey (2005). On the limits of infants' quantification of small object arrays. *Cognition 97*(3), 295–313.

Feigenson, L., S. Carey, and E. Spelke (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology 44*(1), 33–66.

Feigenson, L. and J. Halberda (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences 105*(29), 9926–9930.

Feiman, R. and J. Snedeker (2016). The logic in language: How all quantifiers are alike, but each quantifier is different. *Cognitive Psychology 87*, 29–52.

Flor, E., N. Haslinger, H. Koopman, E. Rosina, M. Roszkowski, and V. Schmitt (2017). Cross-linguistic evidence for a non-distributive lexical meaning of conjunction. In A. Cremers, T. van Gessel, and F. Roelofsen (Eds.), *Proceedings of the 21st Amsterdam Colloquium*, pp. 255–264.

Fodor, J. A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, J. D. (1970). *The Linguistic Description of Opaque Contexts*. Ph. D. thesis, MIT, Cambridge, MA.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland and P. Stateva (Eds.), *Presupposition and Implicature in Compositional Semantics*, pp. 71–120. New York: Palgrave Macmillan.

Frazier, L., J. M. Pacht, and K. Rayner (1999). Taking on semantic commitments, II: collective versus distributive readings. *Cognition 70*(1), 87–104.

Freeman, N. H. and J. A. Stedmon (1986). How children deal with natural language quantification. In I. Kurcz, G. Shugar, and J. Danks (Eds.), *Knowledge and Language*, Volume 39 of *Advances in Psychology*, pp. 21–48. Amsterdam: North-Holland.

Fulkerson, A. L. and R. A. Haaf (2006). Does object naming aid 12-month-olds' formation of novel object categories? *First Language 26*(4), 347–361.

Furrer, S. D. and B. A. Younger (2005). Beyond the distributional input? A developmental investigation of asymmetry in infants' categorization of cats and dogs. *Developmental Science 8*(6), 544–550.

Gajewski, J. (2005). *Neg-Raising: Polarity and Presupposition*. Ph. D. thesis, MIT, Cambridge, MA.

Gelman, S. A., S.-J. Leslie, R. Gelman, and A. Leslie (2019). Do children recall numbers as generic? A strong test of the generics-as-default hypothesis. *Language Learning and Development 15*(3), 217–231.

Gergely, G. and G. Csibra (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences 7*(7), 287–292.

Gergely, G., Z. Nádasdy, G. Csibra, and S. Bíró (1995). Taking the intentional stance at 12 months of age. *Cognition 56*(2), 165–193.

Geurts, B. (2003). Quantifying kids. *Language Acquisition 11*(4), 197–218.

Gil, D. (1995). Universal quantifiers and distributivity. In E. Bach, E. Jelinek, A. Kratzer, and B. H. Partee (Eds.), *Quantification in Natural Languages*, pp. 321–362. Dordrecht: Springer Netherlands.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition 1*(1), 3–55.

Gleitman, L. R., K. Cassidy, R. Nappa, A. Papafragou, and J. C. Trueswell (2005). Hard words. *Language Learning and Development 1*(1), 23–64.

Goodman, N. (1955). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and Semantics*, pp. 41–58. New York: Academic Press.

Harbour, D. (2014). Paucity, abundance, and the theory of number. *Language 90*(1), 185–229.

Hartsuiker, R. J. and C. Westenberg (2000). Word order priming in written and spoken sentence production. *Cognition 75*(2), B27–B39.

Haslinger, N. (2022). Non-maximality and vagueness: Revisiting the plural Sorites paradox. In R. Starr, J. Kim, and B. Öney (Eds.), *Proceedings of the Semantics and Linguistic Theory Conference (SALT) 32*, pp. 63–82.

Haslinger, N., A. Hien, E. E. Rosina, V. Schmitt, and V. Wurm (2023). *A unified semantics for distributive and non-distributive universal quantifiers across languages*. Unpublished manuscript. https://lingbuzz.net/lingbuzz/007632.

Haslinger, N., E. Rosina, M. Roszkowski, V. Schmitt, and V. Wurm (2021). Cumulation cross-linguistically. In M. Dočekal and M. Wągiel (Eds.), *Formal Approaches to Number in Slavic and Beyond*, pp. 219–249. Berlin: Language Science Press.

Haslinger, N. and V. Schmitt (2017). Acquisition of semantic type flexibility: The case of conjunction. In C. Mayr and E. Williams (Eds.), *Festschrift für Martin Prinzhorn (= Wiener Linguistische Gazette 82)*, pp. 109–118. Vienna: University of Vienna.

Havron, N., C. Scaff, M. J. Carbajal, T. Linzen, A. Barrault, and A. Christophe (2020). Priming syntactic ambiguity resolution in children and adults. *Language, Cognition and Neuroscience 35*(10), 1445–1455.

Henderson, A. M., Y. Wang, L. E. Matz, and A. L. Woodward (2013). Active experience shapes 10-month-old infants' understanding of collaborative goals. *Infancy 18*(1), 10–39.

Henderson, A. M. and A. L. Woodward (2011). "Let's work together": What do infants understand about collaborative goals? *Cognition 121*(1), 12–21.

Hochmann, J.-R. (2022). Representations of abstract relations in infancy. *Open Mind 6*, 291–310.

Hochmann, J.-R., S. Benavides-Varela, A. Fló, M. Nespor, and J. Mehler (2018). Bias for vocalic over consonantal information in 6-month-olds. *Infancy 23*(1), 136–151.

Hochmann, J.-R., S. Benavides-Varela, M. Nespor, and J. Mehler (2011). Consonants and vowels: Different roles in early language acquisition. *Developmental Science 14*(6), 1445–1458.

Hochmann, J.-R., S. Carey, and J. Mehler (2018). Infants learn a rule predicated on the relation same but fail to simultaneously learn a rule predicated on the relation different. *Cognition 177*, 49–57.

Hochmann, J.-R., S. Mody, and S. Carey (2016). Infants' representations of same and different in match-and non-match-to-sample. *Cognitive Psychology 86*, 87–111.

Huang, Y. T., E. Spelke, and J. Snedeker (2013). What exactly do numbers mean? *Language Learning and Development 9*(2), 105–129.

Hurewitz, F., A. Papafragou, L. Gleitman, and R. Gelman (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development 2*(2), 77–96.

Inhelder, B. and J. Piaget (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. London: Routledge.

Inhelder, B. and J. Piaget (1964). *The Early Growth of Logic in the Child*. London: Routledge.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language 59*(4), 434–446.

Jiang, Y., S. Ren, Y. Shen, Y. Sudo, and R. Breheny (2022). Plural priming revisited: inverse preference and spillover effects. In M. Degano, T. Roberts, G. Sbardolini, and M. Schouwstra (Eds.), *Proceedings of the 23rd Amsterdam Colloquium*, pp. 152–158.

Kamewari, K., M. Kato, T. Kanda, H. Ishiguro, and K. Hiraki (2005). Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion. *Cognitive Development 20*(2), 303–320.

Kamp, H. and B. Partee (1995). Prototype theory and compositionality. *Cognition 57*(2), 129–191.

Karmiloff-Smith, A. (1979). *A Functional Approach to Child Language: A Study of Determiners and Reference*. Cambridge: Cambridge University Press.

Köpcke, K.-M. (1998). The acquisition of plural marking in English and German revisited: Schemata versus rules. *Journal of Child Language 25*(2), 293–319.

Kouider, S., J. Halberda, J. Wood, and S. Carey (2006). Acquisition of english number marking: The singular-plural distinction. *Language Learning and Development 2*(1), 1–25.

Kovács, A. M. (2014). Extracting regularities from noise: Do infants encode patterns based on same and different relations? *Language Learning 64*(2), 65–85.

Kovács, Á. M. and J. Mehler (2009). Flexible learning of multiple speech structures in bilingual infants. *Science 325*(5940), 611–612.

Kovács, A. M., E. Téglás, G. Gergely, and G. Csibra (2017). Seeing behind the surface: communicative demonstration boosts category disambiguation in 12-month-olds. *Developmental Science 20*(6), e12485.

Krifka, M. (1986). *Nominalreferenz und Zeitkonstitution: Zur Semantik von Massentermen, Pluraltermen und Aspektklassen*. Ph. D. thesis, University of Munich, Munich.

Krifka, M. (1990). Boolean and non-Boolean 'and'. In L. Kálmán and L. Pólos (Eds.), *Papers from the Second Symposium on Logic and Language*, Budapest, pp. 161–188. Akadémiai Kiadó.

Križ, M. (2015). *Aspects of Homogeneity in the Semantics of Natural Language*. Ph. D. thesis, University of Vienna, Vienna.

Križ, M. (2016). Homogeneity, non-maximality and *all*. *Journal of Semantics 33*(3), 493–539.

Križ, M. and E. Chemla (2015). Two methods to find truth-value gaps and their application to the projection problem of homogeneity. *Natural Language Semantics 23*(3), 205–248.

Križ, M. and B. Spector (2021). Interpreting plural predication: Homogeneity and non-maximality. *Linguistics and Philosophy 44*(5), 1131–1178.

Landman, F. (1989). Groups I. *Linguistics and Philosophy 12*, 559–605.

Lasersohn, P. (1998). Generalized distributivity operators. *Linguistics and Philosophy 21*(1), 83–93.

Lasersohn, P. (1999). Pragmatic halos. *Language 3*(75), 522–551.

Le Corre, M. and S. Carey (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition 105*(2), 395–438.

Leahy, B. P. and S. E. Carey (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences 24*(1), 65–78.

Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bäuerle, C. Schwarze, and A. v. Stechow (Eds.), *Meaning, Use and Interpretation of Language*, pp. 302–323. Berlin: DeGruyter.

Link, G. (1987). Generalized quantifiers and plurals. In P. Gärdenfors (Ed.), *Generalized Quantifiers: Lingustic and Logical Approaches*, pp. 151–180. Dordrecht: Springer.

Löbner, S. (1987). The conceptual nature of natural language quantification. In I. Rusza and A. Szabolcsi (Eds.), *Proceedings of the '87 Debrecen Symposion on Logic and Language*, pp. 81–94. Budapest: Akadémiai Kiadó.

Löbner, S. (2000). Polarity in natural language: Predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy 23*, 213–308.

Macnamara, J. (1986). *A Border Dispute: The Place of Logic in Psychology*. Cambridge, MA: MIT Press.

Magri, G. (2012). No need for a dedicated theory of the distribution of readings of English bare plurals. In A. Chereches (Ed.), *Proceedings of the Semantics and Linguistic Theory Conference (SALT) 16*, pp. 383–402.

Magri, G. (2014). An account for the homogeneity effect triggered by plural definites and conjunction based on double strengthening. In S. P. Reda (Ed.), *Pragmatics, Semantics and the Case of Scalar Implicatures*, Palgrave Studies in Pragmatics, Language and Cognition, pp. 99–145. London: Palgrave Macmillan UK.

Mahowald, K., A. James, R. Futrell, and E. Gibson (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language 91*, 5–27.

Malamud, S. A. (2012). The meaning of plural definites: A decision-theoretic approach. *Semantics and Pragmatics 5*(3), 1–58.

Maldonado, M., E. Chemla, and B. Spector (2017). Priming plural ambiguities. *Journal of Memory and Language 95*, 89–101.

Maldonado, M., E. Chemla, and B. Spector (2019). Revealing abstract semantic mechanisms through priming: The distributive/collective contrast. *Cognition 182*, 171–176.

McCrink, K. and K. Wynn (2004). Large-number addition and subtraction by 9-month-old infants. *Psychological Science 15*(11), 776–781.

McMurray, B. and R. N. Aslin (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy 6*(2), 203–229.

Moyer, M., Z. Husnain, and K. Syrett (2019). Won't somebody think of the children? Beyond maximality with plural definite descriptions. In M. M. Brown and B. Dailey (Eds.), *Proceedings of the 43rd Annual Boston University Conference on Language Development (BUCLD)*, pp. 441–453. Somerville, MA: Cascadilla Press.

Munn, A., K. Miller, and C. Schmitt (2006). Maximality and plurality in children's interpretation of definites. In D. Bamman, T. Magnitskaia, and C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development (BUCLD)*, pp. 377–387. Somerville, MA: Cascadilla Press.

Musolino, J. (2004). The semantics and acquisition of number words: Integrating linguistic and developmental perspectives. *Cognition 93*(1), 1–41.

Musolino, J. (2009). The logical syntax of number words: Theory, acquisition and processing. *Cognition 111*(1), 24–45.

Nouwen, R. (2016). Plurality. In M. Aloni and P. Dekker (Eds.), *The Cambridge Handbook of Formal Semantics*, pp. 267–284. Cambridge: Cambridge University Press.

Pagliarini, E., G. Fiorin, and J. Dotlačil (2012). The acquisition of distributivity in pluralities. In A. K. Biller, E. Y. Chung, and A. E. Kimball (Eds.), *Proceedings of the 36th Annual Boston University Conference on Language Development (BUCLD)*, pp. 387–399. Somerville, MA: Cascadilla Press.

Paillé, M. (2022). Subatomic homogeneity without the excluded middle presupposition: An argument from conjunction. In D. Gutzmann and S. Repp (Eds.), *Proceedings of Sinn und Bedeutung 26*, pp. 658–676.

Park, T.-Z. (1978). Plurals in child speech. *Journal of Child Language 5*(2), 237–250.

Partee, B. H. (2011). Formal semantics: Origins, issues, early impact. In B. H. Partee, M. Glanzberg, and J. Skilters (Eds.), *Formal Semantics and Pragmatics. Discourse, Context, and Models. The Baltic Yearbook of Cognition, Logic and Communication*, Volume 6, pp. 1–52. Manhattan, Kansas: New Prairie Press.

Pauen, S. (2002). Evidence for knowledge-based category discrimination in infancy. *Child Development 73*(4), 1016–1033.

Peirce, J., J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods 51*(1), 195–203.

Philip, W. (1995). *Event Quantification in the Acquisition of Universal Quantification*. Ph. D. thesis, University of Massachusetts, Amherst, MA.

Philip, W. (2011). Acquiring knowledge of universal quantification. In J. de Villiers and T. Roeper (Eds.), *Handbook of Generative Approaches to Language Acquisition*, pp. 351–394. New York, NY: Springer.

Pickering, M. J. and H. P. Branigan (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language 39*(4), 633–651.

Poortman, E. B. (2014). Between intersective and 'split' interpretations of predicate conjunction: The role of typicality. In J. Degen, M. Franke, and N. Goodman (Eds.), *Proceedings of the Formal & Experimental Pragmatics Workshop*, pp. 36–42.

Pun, A., S. A. Birch, and A. S. Baron (2016). Infants use relative numerical group size to infer social dominance. *Proceedings of the National Academy of Sciences 113*(9), 2376–2381.

Quine, W. v. O. (1960). *Word and Object*. Cambridge, MA: MIT Press.

Quinn, P., E. Brown, and R. Hayes (2001). Developmental change in form categorization in early infancy. *British Journal of Developmental Psychology 19*(2), 207–218.

Quinn, P. C. and P. D. Eimas (1998). Evidence for a global categorical representation of humans by young infants. *Journal of Experimental Child Psychology 69*(3), 151–174.

R Core Team (2021). *R: A Language and Environment for Statistical Computing* (computer software). R Foundation for Statistical Computing. https://www.R-project.org/.

Raffray, C. N. and M. J. Pickering (2010). How do people construct logical form during language comprehension? *Psychological Science 21*(8), 1090–1097.

Rasin, E. and A. Aravind (2021). The nature of the semantic stimulus: The acquisition of *every* as a case study. *Natural Language Semantics 29*(2), 339–375.

Roberts, C. (1987). *Modal Subordination, Anaphora and Distributivity*. Ph. D. thesis, University of Massachusetts, Amherst, MA.

Roeper, T. and J. de Villiers (1993). The emergence of bound variable structures. In E. Reuland and W. Abraham (Eds.), *Knowledge and Language*, pp. 105–139. Dordrecht: Springer.

Roeper, Thomas, U. S. and B. Z. Pearson (2004). The acquisition path of quantifiers: Two kinds of spreading. In T. Heizmann (Ed.), *Current Issues in Language Acquisition, UMOP 34*, pp. 97–128. Amherst, MA: Graduate Linguistics Student Association (GLSA).

Rousselle, L. and L. Vossius (2021). Acquiring the cardinal knowledge of number words: A conceptual replication. *Journal of Numerical Cognition 7*(3), 411–434.

Russell, B. (1923). Vagueness. *Australasian Journal of Philosophy and Psychology 1*, 84–92.

Saponaro, C., D. Carioti, M. Riva, and M. T. Guasti (2024). *Language as a mirror of the mind: What can children tell us about distributivity?* Unpublished manuscript. https://doi.org/10.31219/osf.io/vywcs.

Sarnecka, B. W., V. G. Kamenskaya, Y. Yamana, T. Ogura, and Y. B. Yudovina (2007). From grammatical number to exact numbers: Early meanings of 'one', 'two', and 'three' in English, Russian, and Japanese. *Cognitive Psychology 55*(2), 136–168.

Sauerland, U., J. Anderssen, and K. Yatsushiro (2005). The plural is semantically unmarked. In S. Kepser and M. Reis (Eds.), *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*, pp. 413–434. Berlin: Mouton de Gruyter.

Scha, R. (1981). Distributive, collective and cumulative quantification. In J. Groenendijk, T. Janssen, and M. Stokhof (Eds.), *Formal Methods in the Study of Language*, Volume 2, pp. 483–512. Amsterdam: Mathematical Centre Tracts.

Schmitt, V. (2013). *More pluralities*. Ph. D. thesis, University of Vienna, Vienna.

Schwarz, F. (2013). Maximality and definite plurals - experimental evidence. In E. Chemla, V. Homer, and G. Winterstein (Eds.), *Proceedings of Sinn und Bedeutung 17*, pp. 509–526.

Schwarzschild, R. (1994). Plurals, presuppositions and the sources of distributivity. *Natural Language Semantics 2*(3), 201–248.

Schwarzschild, R. (1996). *Pluralities*. Dordrecht: Kluwer Academic.

Shipley, E. F. and B. Shepperson (1990). Countable entities: Developmental changes. *Cognition 34*(2), 109–136.

Shukla, M., K. S. White, and R. N. Aslin (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences 108*(15), 6038–6043.

Simon, T. J., S. J. Hespos, and P. Rochat (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive Development 10*(2), 253–269.

Simon-Pearson, L. and K. Syrett (2018). Assessing truth and speaker knowledge when utterances are not maximally true. In A. B. Bertolini and M. J. Kaplan (Eds.), *Proceedings of the 42nd Annual Boston University Conference on Language Development (BUCLD)*, pp. 708–721. Somerville, MA: Cascadilla Press.

Sophian, C. and C. Kailihiwa (1998). Units of counting: Developmental changes. *Cognitive Development 13*(4), 561–585.

Spector, B. (2016). Multivalent semantics for vagueness and presupposition. *Topoi 35*(1), 45–55.

Sternefeld, W. (1998). Reciprocity and cumulative predication. *Natural Language Semantics 6*(3), 303–337.

Syrett, K. (2014). Mapping properties to individuals in language acquisition. In E. Grillo and K. Jepson (Eds.), *Proceedings of the 39th Annual Boston University Conference on Language Development (BUCLD)*, pp. 398–410. Somerville, MA: Cascadilla Press.

Syrett, K. and A. Aravind (2022). Context sensitivity and the semantics of count nouns in the evaluation of partial objects by children and adults. *Journal of Child Language 49*(2), 239–265.

Syrett, K., C. Kennedy, and J. Lidz (2010). Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics 27*(1), 1–35.

Syrett, K. and J. Musolino (2013). Collectivity, distributivity, and the interpretation of plural numerical expressions in child and adult language. *Language Acquisition 20*(4), 259–291.

Syrett, K. and J. Musolino (2016). All together now: Disentangling semantics and pragmatics with *together* in child and adult language. *Language Acquisition 23*(2), 175–197.

Szagun, G. (2001). Learning different regularities: The acquisition of noun plurals by German-speaking children. *First Language 21*(62), 109–141.

Takahashi, M. (1991). Children's interpretation of sentences containing *every*. In J. Allwood and E. B. Plunkett (Eds.), *University of Massachusetts Occasional Papers in Linguistics 17*, pp. 303–327. Amherst, MA: Graduate Linguistics Student Association (GLSA).

Téglás, E. and L. L. Bonatti (2009). *Evidence for logical representations in 10-month-old infants*. Poster presented at the Society for Research in Child Development (SRCD) Meeting, Denver, Colorado, April 2-4.

Téglás, E., E. Vul, V. Girotto, M. Gonzalez, J. B. Tenenbaum, and L. L. Bonatti (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science 332*(6033), 1054–1059.

Tieu, L., M. Križ, and E. Chemla (2019). Children's acquisition of homogeneity in plural definite descriptions. *Frontiers in Psychology 10*, 2329.

Vendler, Z. (1962). *Each* and *every*, *any* and *all*. *Mind 71*(282), 145–160.

Vizmathy, L., K. Begus, G. Knoblich, G. Gergely, and A. Curioni (2024). Better together: 14-month-old infants expect agents to cooperate. *Open Mind 8*, 1–16.

von Fintel, K. and L. Matthewson (2008). Universals in semantics. *The Linguistic Review 25*(1-2), 139–201.

Warneken, F. and M. Tomasello (2007). Helping and cooperation at 14 months of age. *Infancy 11*(3), 271–294.

Waxman, S. R. and D. B. Markow (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. *Cognitive Psychology 29*(3), 257–302.

Winter, Y. (2001). Plural predication and the Strongest Meaning Hypothesis. *Journal of Semantics 4*(18), 333–365.

Wood, J. N., S. Kouider, and S. Carey (2009). Acquisition of singular-plural morphology. *Developmental Psychology 45*(1), 202–206.

Wynn, K. (1992a). Addition and subtraction by human infants. *Nature 358*, 749–750.

Wynn, K. (1992b). Children's acquisition of the number words and the counting system. *Cognitive Psychology 24*(2), 220–251.

Wynn, K., P. Bloom, and W.-C. Chiang (2002). Enumeration of collective entities by 5-month-old infants. *Cognition 83*(3), B55–B62.

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review 126*(6), 841–864.

Xu, F. and E. S. Spelke (2000). Large number discrimination in 6-month-old infants. *Cognition 74*(1), B1–B11.

Yoon, Y. (1996). Total and partial predicates and the weak and strong interpretations. *Natural Language Semantics 4*(3), 217–236.