Testing for Dual-Task Effects on Spreading

Chapter 6
Abstract

Children’s non-adult-like distributivity interpretations have been linked to another phenomenon in quantifier acquisition, namely children’s spreading errors. Spreading is a well-studied error in children’s interpretation of universal quantifiers. Children up to the age of 10 will incorrectly reject sentences like ‘each girl is building a sandcastle’ when they see a picture in which there is an extra sandcastle. They refer to this extra object to explain their rejection. Musolino (2009) argues that children’s distributivity interpretations might be related to their spreading errors and that both phenomena could share the same origin. He based his theory on a processing model developed by Geurts (2003). Geurts (2003) argues that children’s spreading errors are a result of their processing limitations. He claims that children up to 10 years old misapply the interpretation of weak quantifiers to strong quantifiers, for the simple reason that weak quantifiers are easier to process. This resource-based theory entails that children’s spreading errors should disappear when their working memory capacity increases. For adults, this entails that a limited working memory capacity might lead to spreading. Recently, using a resource-demanding task, Bott and Schlotterbeck (2018) seem to find spreading in adults, serving as evidence for Geurts’ resource-based account. In the present study, we further investigated this account by limiting adults’ working memory capacity in a dual task. Our adult participants made no spreading errors. Our results, therefore, conflict with Bott and Schlotterbeck (2018), but are consistent with other experimental results with children that found no relationship between children’s working memory capacity and their spreading errors (de Koster et al., 2018). Besides serving as evidence against a resource-based explanation of spreading, the results of the current study also serve as evidence against Musolino’s (2009) account that links children’s distributivity interpretations to their spreading errors.
1 Introduction

Children’s distributivity interpretations have been found to differ extensively from the interpretations of adults. For adults it has been found that they accept the distributive interpretation in different degrees for different groups of pluralities (Dotlačil, 2010). Adults prefer a distributive interpretation for distributive quantifiers like *each* (Brooks & Braine, 1996; Syrett & Musolino, 2013). For counting quantifiers like *all* and *both*, on the other hand, adults will accept both a distributive and a collective interpretation depending on context (Brooks & Braine, 1996; Kaup et al., 2002). For group-denoting plurals such as definite plurals and numerals, the distributive interpretation is degraded for adults (de Koster et al., 2017; Frazier et al., 1999; Gil, 1982; Padilla-Reyes, 2018; Pagliarini et al., 2012) and they will prefer a collective interpretation. Unlike adults, children seem to be insensitive to these different types of pluralities. They accept both interpretations to the same degree, sometimes preferring the distributive interpretation (Brooks & Braine, 1996; de Koster et al., 2017; Padilla-Reyes, 2018; Pagliarini et al., 2012; Syrett & Musolino, 2013). The question remains how these differences between adults and children arise. What causes children to be insensitive to the different types of pluralities?

One account that tries to explain the non-adult-like distributivity interpretations of children is the processing account by Musolino (2009). Musolino (2009) argues that children’s non-adult like distributivity interpretations might be linked to another well-known phenomenon in the development of quantification: spreading errors. He does this by extending an existing account of spreading developed by Geurts (2003) and argues that this account might also explain children’s non-adult-like distributivity interpretations. Geurts’ (2003) account is resource-based and attributes children’s spreading errors to their processing limitations. We previously discussed and examined this account in Chapter 4, where we presented a study that examined the relationship between children’s working memory capacity and their spreading errors (de Koster et al., 2018). We found evidence against such a relationship: children’s spreading errors did not correlate with their working memory capacity.

The work presented in this chapter will examine Geurts’ (2003) account further, this time not by focusing on children, but on adults. Using a dual-task experiment we try to find out whether or not it is possible to induce spreading errors in adults, by limiting their working memory capacity. The results of this study will not only shed light on the origin of children’s spreading errors, but will also provide more insight into the unified processing account proposed by Musolino (2009). Both Geurts’ (2003) and Musolino’s (2009) accounts will be discussed in more detail later. We will start, however, with a short explanation of children’s spreading errors and an overview of the different accounts of spreading.
2 Children’s Spreading Errors

Quantifiers describe a relation between sets (Barwise and Cooper, 1981). Consider sentence (1) with the universal quantifier *each*.

(1) Each girl is building a sandcastle.

In sentence (1) the quantifier *each* denotes a relation between the set of girls and the set of sandcastle builders. The set of girls, the restrictor set, is provided by the subject NP. The set of sandcastle builders, the scopal set, is provided by the predicate. The relationship between these two sets is determined by the quantifier. Since the quantifier *each* is a universal quantifier it denotes a subset relation, meaning that the set of girls must be a subset of the set of sandcastle builders for the sentence to be true. In other words, for all girls it must be true that they are building a sandcastle. One girl that is not building a sandcastle is enough to make sentence (1) an incorrect description. The number of sandcastles, however, is irrelevant to the truth value and both Figure 1 and Figure 2 can be said to depict (1).

![Figure 1. Extra Object](image1)

![Figure 2. 1-to-1 Pairing](image2)

Inhelder and Piaget (1964) were the first to report persistent errors in children’s understanding of quantifiers. Children up to the age of 10, will often incorrectly reject a sentence like (1) describing a scene such as presented in Figure 1. This error is called a spreading error. Children base their rejection on the extra sandcastle, letting the restrictor set ‘spread’ to the scopal set. All members of the scopal set, including the extra sandcastle, are incorrectly considered relevant by children. The extra sandcastle is actually irrelevant for determining the truth value of sentence (1), due to the conservativity property of universal quantifiers. In contrast, children behave like adults when they are presented with a scene such as in Figure 2 in combination with sentence (1). In this case, they will correctly accept sentence (1) as a description of the 1-to-1 pairing situation depicted in Figure 2.
Children’s spreading errors are pervasive and have been extensively studied (among others: Crain et al., 1996; Drozd, 2001; Drozd & van Loosbroek, 1999; Inhelder & Piaget, 1964; Philip, 1995; Roeper et al., 2004; Roeper & de Villiers, 1993). They occur in many languages and across different universal quantifiers. Spreading errors start around age five, seem to peak at age seven (Aravind et al., 2017; Roeper et al., 2004) and persist until quite late (around age ten). Despite the numerous studies that have been dedicated to spreading, its cause and how children overcome the error is still subject to major debate, with multiple competing theoretical accounts.

As mentioned in the introduction, the current study focuses on Geurts’ account of spreading (2003), according to which spreading errors are caused by processing limitations. This account argues that when these processing limitations disappear due to cognitive development, children’s spreading errors will resolve. A study by de Koster et al. (2018) (presented in Chapter 4) tested this account, and examined whether or not children’s spreading errors were related to their working memory capacity. Their results showed no effect of working memory capacity on children’s spreading errors, serving as evidence against a resource-based account of spreading.

In a recent study, however, Bott and Schlotterbeck (2018) claimed to have found evidence supporting a resource-based account of spreading. They induced spreading errors in adults by letting them perform a cognitively demanding task, limiting the working memory capacity of the adult participants and thereby simulating children’s underdeveloped working memory capacity.

These results appear to contradict the results from de Koster et al. (2018) presented in Chapter 4 and ask for a re-examination. We therefore tested adults in a dual-task setting, limiting their working memory capacity, using similar materials as used by de Koster et al. (2018). If children’s spreading errors are indeed caused by an underdeveloped working memory capacity, we would expect to see spreading errors in adults as well when their working memory is limited. However, if children’s spreading errors are not related to their cognitive resources, as was concluded by de Koster et al. (2018), we will not see spreading errors in adults irrespective of their working memory limitations. The latter will serve as evidence against both Geurts’ (2003) resource-based account of spreading and Musolino’s (2009) extension of this account, in which he proposes a unified explanation of children’s spreading errors and non-adult-like distributivity interpretations. Before we will discuss the current study, we will first give an overview of the different accounts of spreading, since the resource-based account by Geurts (2003) is not the only explanation for spreading errors.
3 Different Accounts of Spreading

Spreading is a widely examined phenomenon. Three main lines of research can be distinguished: knowledge-based accounts, pragmatic accounts and resource-based accounts.

The knowledge-based accounts argue that the representations of universal quantifiers differ between adults and children. Syntactic approaches by Roeper and de Villiers (1993), Roeper et al. (2004) and Philip (1995), for example attribute spreading errors to differences in syntactic representations between adults and children. They argue that children initially treat universal quantifiers as if they are sentential adverbs like *always*. This allows them to quantify over events and interpret (1) as if it was: *a girl is always building a sandcastle*. In this case the restrictor set is determined on the basis of events. For children, sub-events of an event in which a girl is building a sandcastle are also taken into considerations as events. This means that for every event that contains either a girl or a sandcastle or both it must be the case that a girl is building a sandcastle. The extra sandcastle in Figure 1, for example, could be part of an event in which a girl is building a sandcastle and is therefore taken into consideration for assessing the truth value of the sentence. The sandcastle, however, does not correspond to a complete event and the sentence is therefore rejected. Roeper et al. (2004) present a developmental path in which a universal quantifier starts as a sentential adverb and step by step changes into the adult representation of a universal quantifier. They argue that spreading errors are not related to cognition and pragmatics, but rather result from the challenge of grammar construction. Children have to understand and learn that quantifiers belong inside a DP and cannot spread to other DPs.

Pragmatic accounts, on the other hand, attribute spreading errors to flaws in the experimental set up. The reasons why children make spreading errors are purely pragmatic and have to do with the testing situations being pragmatically infelicitous. Crain et al. (1996), for example, claim that children’s spreading errors might arise due to inappropriate experimental materials that do not meet the condition of plausible dissent. In short, this means that a yes-no question should only be asked if both associated outcomes are made salient by presenting situations that could make the question true but also could make the question false. This presentation of both alternatives justifies asking the question. If only one alternative is presented it is infelicitous to ask the question. In typical truth-value judgement tasks, which are very common in spreading research, this condition is not met, since only one picture is presented.
A picture depicting a situation that makes the question false is usually not provided to the child.\(^1\) The child, therefore might wonder why the question is being asked and as a result incorrectly assume that something else must be going on. They may not understand why an adult would ask such an easy question and deduce that something must be going on that justifies asking the question. In such a situation, the child may come up with a reason that justifies the infelicitous question and a possible explanation might be that the extra object is taken into consideration, because of its salience. Crain et al. (1996) showed in several experiments that children who initially made classical spreading errors, did not do so anymore when the condition of plausible dissent was met. They conclude from this that children’s interpretations of universal quantifiers do not differ from adults.

The claims made by this pragmatic account raise the question why adults do not show spreading errors, if children’s spreading errors are not due to a lack of knowledge but just flawed experiments. The pragmatic account attributes the differences between adults and children to children’s inability to cope with pragmatically infelicitous discourses. Crain et al. (1996) argue that in normal circumstances older children and adults are more skilled in accommodating pragmatic infelicities than younger children (up to the age of 10 years old). Dealing with these infelicities goes hand in hand with certain aspects of children’s cognitive development (such as decision making, resisting distraction and selecting and keeping strategies). So this pragmatic account concludes that spreading is not caused by a lack of linguistic knowledge, but rather is the side-product of an inappropriate experimental design.

The third line of research is represented by the resource-based accounts that attribute children’s spreading errors to their underdeveloped cognitive resources. One of these accounts is the weak-strong account (Geurts, 2003), which explains children’s spreading errors as the result of a mismapping from syntax to semantics. It is important to note that this account does not assume that children have incomplete syntactic or semantic knowledge. In essence, the weak-strong account argues that children misapply a simple structure instead of a complex structure, due to cognitive limitations, which ultimately leads to an incorrect interpretations of

\(^1\) Please note that many Truth-Value Judgement Tasks do not only provide correct items. They often also contain items that are incorrect, which require the child to say ‘no’. However, this contrast is between different items. Crain et al. (1996) argue that the condition of plausible dissent, like the condition of plausible assent, should be met within each item. This means that a test item should initially make both outcomes plausible (a ‘yes’ answer meeting the condition of plausible assent and a ‘no’ answer meeting the condition of plausible dissent), so the child can take both outcomes in consideration. See for example the story about the skiers used in Crain et al.’s experiment (1996, page 126) based on the Truth-Value Judgement methodology from Crain and McKee (1985).
quantifiers. Geurts (2003) argues that children treat strong quantifiers like *each* and *all* as if they are weak.\(^2\)

The semantic literature on quantification assumes that there is a basic distinction\(^3\) between strong and weak quantifiers: some quantifiers are allowed in existential *there*-sentences and other are not. Strong quantifiers such as *each* and *all* cannot occur in existential *there*-sentences (2). Weak quantifiers like *some*, on the other hand, can occur in such sentences (3).

(2) *There are *each/all* girls building a sandcastle.*  
(3) *There are some girls building a sandcastle.*

The ability to occur in *there*-sentences is not the only distinction between the two types of quantifiers. From a semantic point of view, the crucial factor is that weak quantifiers are intersective and strong quantifiers are not. Take for example sentence (4).

(4) *Some girls are building a sandcastle.*  
(5) *Each girl is building a sandcastle.*

This sentence can be verified by simply checking that the intersection between the set of girls (the restrictor set) and the set of sandcastle builders (the scopal set) is not empty. This entails that to determine the truth value of a sentence with a weak quantifier, only one set needs to be checked, namely the intersection of the two sets. The weak quantifier *some* does not require a relational interpretation.\(^4\) Only the set of girls who are sandcastle builders needs to be evaluated and any other girls can be ignored. See the left side of Figure 3 for a Venn diagram illustrating this intersective character of the weak quantifier *some*.

A strong quantifier like *each*, on the other hand, does require a relational interpretation and depends on two sets. To determine the truth value of sentence (5), containing the strong quantifier *each*, we have to check if the set of girls that are not ‘sandcastle builders’ is empty or not. More specifically, we need to check the set of girls minus the intersection of the set of girls and the set of sandcastle builders,

---

\(^2\) A similar explanation is given by Drozd (2001). Drozd (2001), however, does not share Geurts’ (2003) view that children’s weak/strong mix-up results from a mismapping from syntax to semantics. A comparison between the two accounts is provided by Geurts (2003, pages 212-213).

\(^3\) This distinction between quantifiers is dubbed the weak-strong distinction by Milsark (1976).

\(^4\) Note that weak quantifiers can call for a relational interpretation in cases where they trigger a domain presupposition.
and check if this subset is empty or not. For strong quantifiers, contrary to weak quantifiers, it is not sufficient to look exclusively at the intersection. The strong quantifier *each* is therefore not intersective but relational between the set of girls that are sandcastle builders (the intersection) and the set of all the girls. The relational character of the strong quantifier *each* is depicted in the Venn diagram on the right side of Figure 3.

\[
\text{Some}(A,B) = \text{True iff } A \cap B \neq \emptyset \quad \text{Each}(A,B) = \text{True iff } A \cdot (A \cap B) = \emptyset
\]

![Venn diagrams](image)

**Figure 3.** Venn diagrams depicting the weak-strong distinction.

From this analysis it follows that weak quantifiers are simpler than strong quantifiers. According to Geurts (2003), strong quantifiers are harder to interpret, because their semantic representation is more complex. Geurts (2003) describes it as follows: “When processing a sentence of the form ‘Q X are Y”, X and Y must be processed separately if Q is strong, whereas if Q is weak, X and Y can be interpreted as if they were conjoined.” This difference in complexity is illustrated in Figure 4. Geurts (2003) states that this differentiation is not only supported by semantic notions, but by experimental findings as well. It has been found that weak quantifiers are easier to process than strong quantifiers. Just (1974) for example found that sentences with weak quantifiers had shorter response latencies than sentences with strong quantifiers. Similar observations have been made by Meyer (1970).

![Semantic representations](image)

**Figure 4.** Semantic representations of weak versus strong quantifiers (Geurts, 2003).

Crucially, Geurts’ theory (2003) starts with the assumption that children’s knowledge about the grammar of quantification is not deficient. Children know that a strong quantifier requires a more complex structure. Therefore, he attributes the error to a mismapping from grammatical form to semantic representation rather
than a lack of knowledge. Geurts (2003) argues that children treat strong quantifiers as if they were weak, for the simple reason that they are easier to process. He uses representations in Discourse Representation Theory (DRT) to clarify this mismapping from grammatical form to semantic meaning. In a DRT representation there is a position for the quantifier itself (between angle brackets), the nuclear scope and the restrictor (both between square brackets). Example (6) presents the DRT representations for the weak quantifier some (a) and the strong quantifier each (b).

(6) a. Some girls are building a sandcastle.
  \[\text{<some>}[x, y: \text{girl}(x), \text{sandcastle}(y), x \text{ builds } y]\]
b. Each girl is building a sandcastle.
  \[\text{[x: girl}(x)]\text{<each>}[y: \text{sandcastle}(y), x \text{ builds } y]\]

To explain spreading errors, Geurts (2003) argues that children, when interpreting a strong quantifier, start by making a strong construal of a sentence like (7a), since they know that a strong quantifier requires a tripartite structure (consisting of the quantifier, the restrictor and the nuclear scope). This strong construal is presented in (7b). They then misapply a mapping that is consistent with a weak quantifier, as in (6a), because this structure is simply easier. This incorrect representation leaves the restrictor underdetermined, as is presented in (7c).

(7) a. Each girl is building a sandcastle.
  \[\text{[restrictor]}\text{<each>}[\text{nuclear scope}]\]
b. \[\text{[... : ...]}\text{<each>}[x, y: \text{girl}(x), \text{sandcastle}(y). x \text{ builds } y]\]

Geurts (2003) argues that the eventual domain restriction is driven by pragmatic reasons, because the quantifier is not properly restricted. One of the pragmatic constraints on quantifier domain is the fact that quantifiers prefer to have domains that are contextually salient. The child can choose either ‘girls’ or ‘sandcastles’ as the restrictor of the quantifier. If the child considers ‘girls’ are the most salient discourse entity, they will assume it is the intended restrictor of the quantifier. This mapping results in the adult interpretation of the strong quantifier each. If the child, on the other hand, considers ‘sandcastles’ as the most salient discourse entity (for example as a result of the salient extra object presented in Figure 1), and map ‘sandcastles’ as the restrictor of the quantifier, this will evoke a spreading error. This incorrect mapping will lead to an interpretation of sentence (7a) as if it was: each sandcastle is being built by a girl. This interpretation will lead to a spreading error, since one of the sandcastles is not being built by a girl in Figure 1, making the picture-sentence combination false. Geurts’ proposal (2003) is particularly interesting since it can explain the finding that children show spreading errors, but at the same time can also show adult responses. According to Geurts’ account both sets can end up as the restrictor set. Which one eventually is chosen depends on pragmatics. In addition,
Geurts’ proposal (2003) can also explain why children generally point to the extra object as a reason for their incorrect rejection. In short, Geurts’ resource-based account of spreading attributes children’s spreading errors not to a lack of knowledge, but to an incorrect mapping from syntactic form to semantic representation due to processing limitations.

The current section discussed the three main lines of research that try to account for children’s spreading errors. In the remainder of this chapter we will focus on one line of research in particular, the resource-based accounts of spreading. We do this because our aim is to examine a possible relationship between children’s non-adult-like distributivity interpretations and their spreading errors, as suggested by Musolino (2009). In his work, Musolino (2009) argues that the two phenomena might share a similar origin by extending the resource-based weak-strong account by Geurts (2003). The following two sections discuss evidence against and in favor of the resource-based account of spreading and their relationship with children’s distributivity interpretations. We discuss several findings that point in different directions.

3.1 Evidence against resource-based accounts of spreading

Musolino (2009) took Geurts’ (2003) weak-strong account, introduced in the previous section, as a basis and argued that it could be extended to explain children’s distributivity interpretations as well. If the intuition that children’s non-adult distributivity interpretations are related to their spreading errors is indeed correct, the two phenomena are expected to occur and disappear at approximately the same time. Additionally under this account, both phenomena are expected to be related to children’s cognitive resources, since Geurts’ (2003) weak-strong account attributes children’s spreading errors to their processing limitations.

De Koster et al. (2018) tested this proposed relationship between the two phenomena and their relationship with cognitive resources, by examining spreading errors and distributivity preferences in the same children. They compared the weak-strong account to a different account explaining children’s distributivity preferences, namely the implicature account by Dotlačil (2010).5

(8) The girls are building a sandcastle.

Adults prefer a collective interpretation of plural definites like the, such as in (8). For this sentence, they prefer the situation in which multiple girls are building one sandcastle together over a situation in which multiple girls are building their own sandcastles separately. Children on the other hand do not show this preference and accept both distributive and collective interpretations. Dotlačil’s implicature

5 Please note that this chapter refers to the pragmatic account as the implicature account.
account (2010) predicts that this difference is caused by an implicature. In short this means that adults, when interpreting sentences with unmarked plurals such as (8), can reason about a more informative alternative option with each (5). The quantifier each is a marker for distributivity, so its absence points to a collective situation. Young children have an incomplete lexical understanding of the quantifier each, implying that they do not realize its distributive character. They therefore cannot use it as an alternative option to calculate the implicature, resulting in an acceptance of both the distributive and the collective interpretation for sentences like (8).

Both the weak-strong account and the implicature account predict a relationship with children’s cognitive resources. The weak-strong account essentially proposes that children treat strong quantifiers as if they are weak, because weak quantifiers are easier to process due to their less complex semantic structure. Therefore, limitations of working memory or attention will tend to affect the interpretation of strong quantifiers more than they affect the interpretation of weak quantifiers. This entails that children should make fewer spreading errors when they have a higher working memory capacity. The implicature account also makes a connection with cognitive resources. Part of the implicature processing literature has concluded that implicatures are not default inferences. Their calculation is argued to take time, because it occurs after the calculation of the literal (semantic) interpretation (e.g., Bott and Noveck, 2004; Huang and Snedeker, 2009; Bott et al. 2012) and in addition requires memory resources since both alternatives have to be kept active in memory (e.g., De Neys and Schaeken, 2007; Dieussaert et al., 2011; Marty and Chemla, 2013; Marty et al., 2013). These findings suggest that there should be a relationship between children’s developing collective preference and their developing working memory capacity.

De Koster et al. (2018) tested these predictions by letting Dutch four to eleven year old children perform both a working memory word span task, to assess their working memory capacity, and a truth-value judgement task testing both their spreading errors and their distributivity interpretations. The truth-value judgment task tested four picture types (collective, distributive, extra object and extra subject) in combination with two sentence types (the definite plural de ‘the’ (9), and the quantifier elke ‘each’ (10)). The collective and distributive pictures tested children’s distributivity interpretations, whereas the extra object and extra subject pictures tested children’s spreading errors.

---

De Koster et al. (2018) found that children stopped making spreading errors around nine to ten years old, but still did not show the adult collective preference around eleven years old. So the results showed an age gap between children’s development of distributivity and the disappearance of spreading errors, contrary to the prediction that both phenomena should develop around the same time if they are indeed related. Crucial are the findings of the relationship with working memory. De Koster et al. (2018) did not find a relationship between children’s memory scores and production of spreading errors, contrasting with the prediction of the weak-strong account. More importantly, they did find an effect of memory scores on children’s distributivity interpretations. De Koster et al.’s (2018) results therefore suggest that (i) children’s spreading errors and their distributivity preferences have different origins, contrary to Musolino’s account (2009), (ii) that children’s non-adult-like distributivity preferences seem to be the result of the failure to calculate an implicature, and (iii) that spreading is not caused by underdeveloped cognitive resources such as working memory, contrary to the claims of Geurts’ (2003) resource-based account of spreading.

3.2 Evidence in favor of resource-based accounts of spreading

A recent study by Bott and Schlotterbeck (2018) found evidence contradicting the previously discussed findings of de Koster et al. (2018). In short, de Koster et al. (2018) examined Geurts’ resource-based account of spreading (2003), by testing children with a truth-value judgment task and a working memory word span task. Their aim was to find a relationship between children’s spreading errors and their working memory capacity. Their results, however, did not reveal such a relationship.

Bott and Schlotterbeck (2018) also tested the resource-based account of spreading, but took a different approach. They examined the prediction that adults should show spreading errors when their processing limitations are severely limited. They let adults perform a resource-demanding version of a truth-value judgment task, which they argue to be a dual task. We, however, are not completely sure whether or not Bott and Schlotterbeck’s (2018) task is indeed a dual-task. More about this issue can be found in the Discussion section of this chapter (Section 6).
Bott and Schlotterbeck’s (2018) experiment was conducted in German and the adult participants had to incrementally provide truth-value judgments for universally quantified sentences, such as (11), in a self-paced fashion based on a memorized set diagram (see Figure 5).


Each pupil is such that exactly one teacher praised him full of goodwill.

This meant that for each trial participants had to provide truth-value judgments for universally quantified sentences phrase by phrase, while they had to keep a set diagram in mind. Memory load was achieved by showing the set diagram separately from the sentence.

![Diagram](image)

**Figure 5.** Set diagrams used by Bott and Schlotterbeck (2018), with on the left the 1-to-1-Pairing condition and on the right the Extra Object condition.

Participants first had to inspect a set diagram. After this, the diagram disappeared and the sentence (e.g., (11)) was displayed in a self-paced fashion using a moving-window presentation. This means that for each click a new segment appeared at its position in the sentence and the previous segment disappeared. At each segment participants had to decide whether the unfolded sentence still matched the picture or not. They could choose between ‘no, does not match’, which aborted the trial, or ‘yes, go on’, which made the next segment appear. Whenever a true sentence was aborted or a false sentence was accepted, the trial was counted as an error. Bott and Schlotterbeck (2018) tested diagrams that included an extra object and diagrams that did not have an extra object (see Figure 5). The diagrams with the extra object tested spreading errors: the incorrect rejection of the quantified sentence based on the extra object diagram indicated such an error.

---

It has to be noted that Bott and Schlotterbeck (2018) also examined a new type of spreading error which they dubbed branching errors. Since these errors have never been tested in children and are not tested by de Koster et al. (2018), we will not discuss them any further.
Bott and Schlotterbeck’s (2018) results indicated that adults made spreading errors, with a rejection rate of 11.3% for sentences like (11) in combination with the extra object diagram presented in Figure 5. This contrasted with a rejection rate of only 2.3% in an ordinary picture verification task that served as a control task (Experiment 1: Bott & Schlotterbeck, 2018). In addition, they also found differences in response latencies between items showing diagrams with an extra object and without an extra object. Diagrams with extra objects showed an increase in judgment times compared to diagrams without extra objects in both the cognitively demanding task and the ordinary picture verification task. Both the results of the responses and the response latencies seem to point to a resource-based explanation of spreading. The resource-demanding truth-value judgment task induced spreading errors in adults in 11.3% of the cases, in contrast to only 2.3% of spreading errors when the participants did not have a memory load. This seems to indicate that limited cognitive resources result in spreading errors, as predicted by the resource-based account of spreading. The fact that it took more time to judge the diagrams with an extra object is even more evidence for the resource-based accounts.

Besides the differences in language and the focus on either children or adults, the methodologies of the de Koster et al. (2018) study and the Bott and Schlotterbeck (2018) study also seem to be very different. De Koster et al. (2018) examined a possible relationship with cognitive resources by assessing participants’ working memory capacity using a word span task. Bott and Schlotterbeck (2018), on the other hand, took a different approach and did not assess participants’ working memory capacity independently of the linguistic task but limited their working memory capacity using a cognitively demanding truth-value judgment task. Whereas de Koster et al. (2018) used concrete pictures depicting ordinary situations, Bott and Schlotterbeck (2018) used abstract set diagrams. The sentences were also different: de Koster et al. (2018) tested relatively simple transitive sentences that have subject-verb-object order and that are typically used in spreading research. Bott and Schlotterbeck (2018), on the other hand, used more complex sentences.

In light of the seemingly conflicting evidence provided by the studies of de Koster et al. (2018) and Bott and Schlotterbeck (2018), and the methodological differences between the studies, further research is necessary to be able to find out whether the resource-based account of spreading (Geurts, 2003) and the processing account of distributivity (Musolino, 2009) stand any ground. We therefore conducted a dual task experiment using the same materials that have been used by de Koster et al. (2018) to test children, to find out if we can induce spreading errors in adults, similar to the findings of Bott and Schlotterbeck (2018).
4 Method

The procedure of the dual-task experiment presented in the current chapter is similar to the dual-task experiment presented in Chapter 5 (de Koster, Spenader, & Hendriks, 2020a). The dual-task experiment in Chapter 5 tested distributivity interpretations and examined the implicature account by Dotlačil (2010), whereas the current chapter focuses on spreading errors and the resource-based accounts by Musolino (2009) and Geurts (2003). Two of the tested conditions, however, correspond, which makes a comparison between the two studies possible. The conditions will be introduced and discussed in Sections 4.3.1 and 4.5.

4.1 Participants

Forty students from the University of Groningen (10 men; mean age 21.9; age range 18-33) were paid to participate. All participants were native speakers of Dutch. The study was carried out in accordance with the recommendations of the Research Ethics Committee (CETO) of the Faculty of Arts of the University of Groningen and they approved the protocol. We also obtained written informed consent from all participants prior to testing. Contrary to the dual-task experiment presented in Chapter 5, this dual-task experiment did not test a control group that performed the task without a WM load, since a base-line for performance was already established by de Koster et al. (2018) (see Chapter 4).

4.2 Design

The dual-task experiment consisted of two tasks: while participants interpreted a sentence (the linguistic task), we manipulated their WM load by asking them to memorize a sequence of digits (the digit-span task). The experiment had a 2x2x2 design with the factors PICTURE (Extra Object (Figure 1) vs. 1-to-1 Pairing (Figure 2)), SENTENCE (de ‘the’ vs. elke ‘each’) and WM LOAD (low vs. high).

4.2.1 Linguistic Task

The linguistic task was a Sentence-Picture Verification Task. Participants saw a picture on the computer screen and had to judge whether it matched a recorded sentence by pressing a key on the keyboard.

4.2.2 Digit-Span Task

At the start of each trial, before they received the linguistic item, participants had to memorize a sequence of three or six digits (low and high WM load condition, respectively), presented on screen for one second each. Digits were randomly chosen from 1 to 9, and consecutive digits always differed. After each linguistic item, participants had to recall the digits, by typing them in the same order as they appeared.
4.3 Materials

The experiment consisted of four practice items, 64 test items, 48 filler items (the same items as the implicature control items presented in Chapter 5) and sixteen task control items (straightforwardly true or false items to check for attention, consisting of eight true items and eight false items), resulting in a total of 132 items. The experiment was divided in two blocks (preceded by four practice trials), with 64 items per block. The low (three digits) and high (six digits) WM load conditions were presented in different blocks. Block order (either low or high WM load in the first block) was a between-subject factor. The Latin-Square design of the test items together with the factor block order resulted in eight lists. Item order was randomized for each participant and participants were randomly assigned to a list. The practice and task-control items were the same items as the items used in the dual-task experiment presented in Chapter 5 (testing distributivity interpretations).

4.3.1 Target Items

The 64 test items tested the factors PICTURE (Extra Object (Figure 6) vs. 1-to-1 Pairing (Figure 7)) and SENTENCE (de ‘the’ (12) vs. elke ‘each’ (13)).

(12) Elke jongen wast een boot.
    Each boy is washing a boat.

(13) De jongens wassen een boot.
    The boys are washing a boat.

Eight different transitive verbs were used: build, wash, push, pull, carry, lift, hold or paint (in Dutch: bouwen, wassen, trekken, duwen, dragen, tillen, vasthouden, verven) and the grammatical subjects and objects of these verbs varied across items. We used the same pictures and sentences as the de Koster et al. study (2018). Their study, however, consisted of four picture types. For this study we left out the collective pictures and the extra subject pictures, testing only the ‘distributive’ 1-to-1 pairing pictures and the extra object pictures. The de Koster et al. study (2018)
only included 32 test items, so we created 32 more test items to end up with 64 test items in the current experiment. In order to do this, we used the same verbs and grammatical subjects, but with different grammatical objects.

The design resulted in four conditions: Each-Extra Object, Each-1-to-1 Pairing, The-Extra Object and The-1-to-1 Pairing. Each participant saw sixteen items in each of the four conditions, resulting in 64 test items in total (32 items per block). Conditions Each-Extra Object and Each-1-to-1 Pairing tested participants’ spreading errors, whereas conditions The-Extra Object and The-1-to-1 Pairing tested participants’ distributivity interpretations.

### 4.4 Procedure

The participants performed the experiment in a quiet room at the University of Groningen. They sat in front of a computer screen. The pictures were shown on the computer screen and the sentences were played via a wired speaker. The experimenter was present in the room during the entire experiment.

The experiment started with instructions and four practice trials. For each trial, participants first saw a digit sequence on screen, followed by a picture and a recorded sentence. The recorded sentence was played only once. They then had to judge the acceptability of the sentence by pressing a green (accept) or red (reject) key. Finally, they had to type in the memorized digits. Participants had ten seconds to judge the sentences, with a visual warning message after seven seconds. Next, they had five seconds to recall the digits in the low WM load condition and ten seconds in the high WM load condition. Pilot testing had shown that this provided participants with sufficient time.

Each trial ended with feedback to the participant on how many digits were recalled correctly. A waiting penalty (cf. van Rij et al., 2013) ensured that participants focused on the WM task, and prevented rushing: one incorrect digit resulted in a one-second waiting penalty, two incorrect digits in a two-second waiting penalty, etcetera. Self-paced breaks were provided after every sixteen items and participants had a forced break of at least two minutes in between the two blocks. This procedure is the same as the procedure of the dual-task experiment presented in Chapter 5.

Per trial the following three measures were collected: accuracy in reproducing the digits in the digit-span task, and yes/no responses and response times for the test items and task control items in the linguistic task. The filler items were identical to the implicature control items used in Chapter 5 and will not be discussed any further in this chapter, since the aim of this chapter is not to examine implicatures, but to test spreading errors under WM load and to examine a possible relationship between spreading errors and distributivity interpretations.
4.5 Predictions

Conditions Each-Extra Object and Each-1-to-1 Pairing examined spreading, and can be compared to the standard spreading conditions tested by Bott and Schlotterbeck (2018).

Condition Each-Extra Object was designed to assess spreading errors. A ‘no’ response (rejection) in this condition indicates a spreading error. If it is indeed the case that a limited WM capacity can induce spreading errors in adults, as is argued by the resource-based account of spreading, condition Each-Extra Object should be affected by WM load, with a higher number of spreading errors under high WM load.

Condition Each-1-to-1 Pairing tested the symmetric situation without an extra object, similar to the condition without an extra object by Bott and Schlotterbeck (2018). This condition is expected to be unaffected by WM load, in line with the results of Bott and Schlotterbeck (2018). If spreading errors are indeed caused by processing limitations, we should see a difference in response times between conditions Each-Extra Object and Each-1-to-1 Pairing. Following the resource-based account, it should take more time to interpret items of condition Each-Extra Object than items of Each-1-to-1 Pairing. This finding would be similar to the findings of Bott and Schlotterbeck (2018), who found an increase in response times for items with a diagram including an extra object.

Conditions The-1-to-1 Pairing and The-Extra Object tested participants’ distributivity interpretations. It is important to note that condition The-1-to-1 Pairing is equal to condition The-Distributive examined in Chapter 5. A ‘yes’ response (acceptance) to this condition indicates a literal (distributive) interpretation of sentences like (13) and a ‘no’ response (rejection) indicates a pragmatic interpretation of sentences like (13). A pragmatic ‘no’ response requires an implicature and results in a preference for the collective interpretation over the distributive interpretation. Condition The-1-to-1 Pairing is expected to be affected by WM load, since implicature calculation is assumed to require cognitive resources. The results of the dual-task experiment in Chapter 5 confirmed this prediction: condition The-Distributive turned out to be affected by WM load, with fewer collective interpretations under WM load. We therefore expect to see similar results for condition The-1-to-1 Pairing in the current study.

Condition The-Extra Object is a bit harder to interpret. A rejection of items of this condition could either indicate a spreading error for sentences with definite plurals (based on the extra object) or it could indicate a preference for the collective interpretation. The results of this condition, therefore, need to be compared to the results of conditions Each-1-to-1 Pairing and The-1-to-1 Pairing, to be able to properly interpret them. Please note, however, that a relationship with cognitive
resources is predicted for both possibilities. However, if condition The-Extra Object turns out to be unaffected by WM load, this would be evidence against both the implicature account explaining distributivity interpretations and the resource-based account explaining spreading errors (and possibly explaining distributivity interpretations).

5 Results

Two participants were excluded from the analysis, because they gave incorrect answers to more than 25% of the task-control items. The remaining thirty-eight participants were included in the analysis.

5.1 Digit-Span Task

Participants remembered 94% of the digits correctly in the low WM load condition (three digits) and 74% of the digits in the high WM load condition (six digits). This drop in performance is significant (paired-t (37) = 11.651; p < .001), indicating that the high WM load condition was indeed more difficult. Furthermore, linguistic condition had no effect on the percentage correctly recalled digits. This shows that participants focused on digit recall performance throughout the experiment, irrespective of linguistic condition.

5.2 Linguistic Task

5.2.1 Responses to Test Items

Figure 8 shows the mean acceptance rates for all four linguistic conditions per WM load condition. Please note that the No WM Load condition presents the data of de Koster et al.'s study (2018), who performed the same linguistic task without WM load. The collected data were analyzed using generalized mixed-effect logistic modelling (function glmer(): lme4 package in R, version 3.6.3).

Figure 8 reveals that the adult participants made virtually no spreading errors, since the acceptance rate for condition Each-Extra Object is almost at ceiling in both the low (99.3%) and high (99.6%) WM load conditions. These findings are similar to the findings of de Koster et al. (2018), who also did not find spreading errors in adults using a similar task without WM load. They, however, contradict the results of Bott and Schlotterbeck (2018), who did find spreading errors by adults under WM load. Condition Each-1-to-1 Pairing is fully accepted too, regardless of WM load condition. This is as expected and is in line with previous findings (Bott & Schlotterbeck, 2018; de Koster et al., 2018), since the interpretation of sentences in this condition is straightforward and the pictures do not contain an extra object.
Conditions The-Extra Object and The-1-to-1 Pairing show lower acceptance rates of around 73%, but again with no difference between the two WM load conditions. However, these two conditions do show differences with the results of de Koster et al. (2018), who tested participants without a WM load. The acceptance rates of 25% and 27% found by de Koster et al. (2018) on the task without WM load are much lower than the acceptance rates of 73% in the present study. These results are in line with the findings of Chapter 5, where we found a similar effect of WM load on condition The-Distributive.

Figure 8. Mean acceptance rates per linguistic condition per WM load condition. Error bars show standard error. Sentences contain either *elke* ‘each’ or *de* ‘the’ and pictures show either a situation with an extra object or 1-to-1 pairing between grammatical subjects and objects. The figure also presents the adult results of de Koster et al. (2018) who performed the same linguistic task without a WM load.

Table 1. Overview of the final model for the target response analysis, with reference levels: Sentence: ‘Elke’ (Each), Picture: Extra Object, and Block: 1.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate ($\beta$)</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>6.132</td>
<td>0.530</td>
<td>11.574</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Sentence ‘De’ (The)</td>
<td>-2.940</td>
<td>0.872</td>
<td>-3.372</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Picture 1-to-1 Pairing</td>
<td>-0.157</td>
<td>0.435</td>
<td>-0.360</td>
<td>.719</td>
</tr>
<tr>
<td>Block 2</td>
<td>-0.268</td>
<td>0.497</td>
<td>-0.539</td>
<td>.590</td>
</tr>
</tbody>
</table>
The collected data were analyzed using generalized mixed-effect logistic modelling (function glmer(): lme4 package in R, version 3.6.3). Models were constructed via an iterative forward fitting procedure with model comparisons (cf. Baayen et al., 2008) based on the evaluation of Akaike’s information Criterion (AIC). An AIC decrease of more than two indicates that the goodness of fit of the model improves significantly (Akaike, 2011). We determined whether the following fixed-effect factors improved the goodness of fit of the model: SENTENCE (each, the), PICTURE (extra object, 1-to-1 pairing), WM LOAD (low, high), BLOCK (first, second), BLOCK ORDER (low first, high first), and VERB. The dependent variable was the response (0 for rejection, 1 for acceptance). The final model (Table 1) included the fixed factors SENTENCE, PICTURE, and BLOCK. The factors WM LOAD, BLOCK ORDER and VERB did not significantly improve the model fit and were left out. The maximal random-effects structure licensed by the data included a random intercept for participants and by-participant random slopes for sentence, picture and block.

The final model only revealed a main effect of SENTENCE ($\beta = -2.940; z = -3.372; p < .001$), indicating lower acceptance rates for sentences with de ‘the’ compared to sentences with elke ‘each’. This also becomes clear from Figure 6, which shows that the conditions with elke ‘each’ are fully accepted and the conditions with de ‘the’ show lower acceptance rates of 73%. An initial significant effect of BLOCK ($\beta = -1.026; z = -4.681; p < .001$), indicating lower acceptance rates in block 2 compared to block 1, disappeared when the factor was added as a random slope for participants ($\beta = -0.268; z = -0.539; p = .590$).

An additional model, collapsing the factors SENTENCE and PICTURE into the factor CONDITION, reveals no significant difference between conditions Each-Extra Object and Each-1-to-1 Pairing ($\beta = -0.556; z = -0.731; p = .465$) and conditions The-Extra Object and The-1-to-1 Pairing ($\beta = 0.025; z = 0.112; p = .911$). Both conditions Each-Extra Object and Each-1-to-1 Pairing do differ significantly from conditions The-Extra Object ($\beta = -7.031; z = -9.638; p < .001$ and $\beta = -6.475; z = -10.409; p < .001$, respectively) and The-1-to-1 Pairing ($\beta = -7.006; z = -9.606; p < .001$ and $\beta = -6.450; z = -10.376; p < .001$, respectively). The finding that the factors WM LOAD and BLOCK ORDER did not improve the model fit, shows that there was no influence of WM load. Participants responded similarly in both WM load conditions and it also did not make a difference whether they received a low WM load or a high WM load first (block order). However, when we compare the results of conditions The-Extra Object and The-1-to-1 Pairing to the No WM load condition (previous results from de Koster et al., 2018), we can see a difference. The acceptance rates of both WM load conditions increased from around 25% to around 73%, pointing towards an effect of WM load. This shows that the dual-task design is sensitive enough to load participants’ WM. The results of conditions The-Extra Object and The-1-to-1 Pairing will be discussed further in the Discussion section (Section 6).
5.2.2 Responses to Task Control Items

The task control items (straightforwardly true or false items) were included to check for participants’ attention to the linguistic task. Overall participants\(^8\) responded to 95% of all task control items correctly, which shows that they paid sufficient attention to the linguistic task.

To investigate whether the absence of spreading errors and the higher acceptance rates for conditions The-Extra Object and The-1-to-1 Pairing could be caused by a general ‘yes’-bias (induced by the dual-task setting), we also analyzed participants’ performance on the false task control items (that required a ‘no’ response). The participants answered 93% of the false control items correctly in the low WM load condition (three digits) and they answered 88% of the false control items correctly in the high WM load condition (six digits). This difference was not significant (paired-t (36) = 1.846; p = .073). Participants gave similar numbers of ‘no’ responses on the false control items regardless of WM load. We can therefore conclude that our results cannot be attributed to a general ‘yes’-bias.

5.2.3 Response Times

Since Bott and Schlotterbeck (2018) analyzed response times (RTs) and found a significant increase in response times in the extra object condition compared to the condition without the extra object (similar to our 1-to-1 Pairing condition), we also analyzed responses times.\(^9\) Figure 9 presents boxplots of the RTs for all four linguistic conditions per WM load condition. RTs were measured from the onset of picture and sentence presentation until button press. Outliers were excluded following the interquartile range rule, excluding data points that are more than one and a half times the interquartile range below the first and above the fourth quartile (6% of the data was removed).

The model for the response time analysis was again constructed via an iterative forward fitting procedure, similar to the procedure for the response analyses. We analyzed the log-transformed RTs using linear mixed-effect modelling (function `lmer()`: lme4 package in R, version 3.6.3). We determined whether the following fixed-effect factors improved the goodness of fit of the model using the Akaike Information Criterion (AIC): `CONDITION`, `WM LOAD` (low, high), `BLOCK` (first, second) and `BLOCK ORDER` (low first, high first). The final model (Table 2) included the fixed factors `CONDITION`, `WM LOAD` and `BLOCK`. We also included random intercepts for participants,\(^8\) Two participants were excluded from the analysis, because they gave incorrect answers to more than 25% of the task control items. The Results therefore only include data of the remaining thirty-eight participants.

\(^9\) Please note that the de Koster et al. study (2018) did not collect response times.
items, and by-participant random slopes for CONDITION and WM LOAD. The dependent variable was the response time in milliseconds (log-transformed).

Figure 9. Boxplots of response times per linguistic condition and WM load condition. Notches indicate 95% confidence intervals (Chambers et al., 1983) and values report medians.

Table 2. Overview of the model analyzing the response times, with reference levels: Condition: Each-Extra Object, WM Load: low and Block: 1.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate (β)</th>
<th>SE</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>7.888</td>
<td>0.020</td>
<td>56</td>
<td>395.034</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Condition Each-1-to-1 Pairing</td>
<td>-0.011</td>
<td>0.012</td>
<td>33</td>
<td>-0.941</td>
<td>.354</td>
</tr>
<tr>
<td>Condition The-Extra Object</td>
<td>0.019</td>
<td>0.013</td>
<td>44</td>
<td>1.397</td>
<td>.169</td>
</tr>
<tr>
<td>Condition The-1-to-1 Pairing</td>
<td>0.015</td>
<td>0.016</td>
<td>33</td>
<td>0.967</td>
<td>.341</td>
</tr>
<tr>
<td>WM Load high</td>
<td>0.020</td>
<td>0.014</td>
<td>36</td>
<td>1.419</td>
<td>.164</td>
</tr>
<tr>
<td>Block 2</td>
<td>-0.031</td>
<td>0.016</td>
<td>64</td>
<td>-1.973</td>
<td>.053</td>
</tr>
</tbody>
</table>

None of the factors turned out to be significant. The final model with condition Each-Extra Object as the reference level showed no difference in RTs between condition Each-Extra Object and condition Each-1-to-1 Pairing ($\beta = -0.011; t = -0.941; p = .354$). There was also no difference between condition Each-Extra Object and the other two linguistic conditions (The-Extra Object: $\beta = 0.019; t = 1.397; p = .169$, The-1-to-1 Pairing: $\beta = 0.015; t = 0.967; p = .341$). Relevelling the model with condition Each-1-to-1 Pairing as the reference level shows that conditions The-Extra Object
and The-1-to-1 Pairing both had higher RTs than condition Each-1-to-1 Pairing (The-Extra Object: $\beta = 0.030; t = 2.529; p < .05$, The-1-to-1 Pairing: $\beta = 0.027; t = 2.033; p < .05$), in contrast to condition Each-Extra Object that showed similar RTs. This finding contradicts the findings of Bott and Schlotterbeck (2018), who found significantly longer RTs for pictures with an extra object compared to pictures without an extra object.

An initial significant effect of WM LOAD ($\beta = 0.018; t = 2.409; p < .05$), indicating higher RTs under high WM load compared to low WM load, disappeared when the factor was added as a random slope for participants ($\beta = 0.020; t = 1.419; p = .161$). The final model did show a trend towards significance for the factor BLOCK ($\beta = -0.031; t = -1.973; p < .1$), indicating that RTs were lower in block 2 compared to block 1. This finding could be explained as an effect of task experience. Similar block effects have been observed in Chapter 5 and are also found in other dual-task studies (e.g., Van Rij et al., 2013).

6 Discussion

In the current study we further investigated Geurts’ (2003) weak-strong account. This account is resource-based and attributes children’s spreading errors not to a lack of knowledge, but to an incorrect mapping from syntactic form to semantic representation caused by processing limitations. Bott and Schlotterbeck’s (2018) results support this account, as they found that a limited working memory capacity in adults resulted in spreading errors. The findings of de Koster et al. (2018), however, are not in line with the weak-strong account and contradict the results of Bott and Schlotterbeck (2018). De Koster et al. (2018) did not find a relationship between children’s working memory capacity and their spreading errors.

Based on these contradicting findings we wanted to find out whether or not it is possible to induce spreading errors in adults by letting them perform a cognitively demanding dual task using the same materials as the study of de Koster et al. (2018). As previously explained, the methodologies of the two studies were very different. Both studies tested different participant groups and used different sentence types and different picture types. Conducting a dual task with adults using the materials of de Koster et al. (2018) was therefore the logical next step as it could shed light on the question whether spreading errors are indeed caused by processing limitations or that the results of Bott and Schlotterbeck (2018) might have been affected by their design.

The results of the current dual task contradict the findings of Bott and Schlotterbeck (2018). We did not find evidence that loading adults’ working memory leads to spreading errors. Both the response results and the response time results were not in line with the findings of Bott and Schlotterbeck (2018). Our adults hardly made
any spreading errors and we did not find a difference in response times between the spreading conditions with and without an extra object (Each-Extra Object and Each-1-to-1 Pairing). In contrast, Bott and Schlotterbeck (2018) found 11.3% spreading errors in their incremental truth-value judgment task that tested adults under working memory load. In addition, Bott and Schlotterbeck’s (2018) response times also differed between items with and without an extra object, which they argue to be caused by the presence of the extra object.

Why did Bott and Schlotterbeck (2018) find adult spreading errors and the present study, based on the previous work of de Koster et al. (2018), did not? We believe that several differences between the experimental designs might explain the differences in results between the two experiments and why the adults in the Bott and Schlotterbeck (2018) study seemed to spread.

The first difference between the two studies that may explain their different results concerns the dual-task design. Bott and Schlotterbeck (2018) claim that their task is a dual task that loads adults' working memory capacity. Their participants performed a resource-demanding version of a truth-value judgement task. The truth-value judgment task was divided in two sub-tasks, of which the second one depends on the first one: memorizing an abstract set diagram and judging a quantified sentence. They enhanced memory load by showing the diagram separately from the sentence.

A standard dual task, however, actually requires two independent tasks, a main task and a secondary task, that need to be performed simultaneously. The secondary task limits the working memory capacity of the participants in such a way that the performance on the main task may be affected. This was not the case in the task of Bott and Schlotterbeck (2018), which consisted of two sub-tasks that depended on each other, rather than two independent tasks. The validation of the sentence was highly dependent on the correct memorization of the diagram. What Bott and Schlotterbeck (2018) actually did is make an ordinary truth-value judgement task harder by presenting the diagram apart from the sentence. The diagram needed to be memorized to be able to correctly validate the sentence. This design choice seems to confuse memory load with correct memorization. Memory load should be similar for each condition, in order to find out whether memory load affects a certain condition in comparison with other conditions. However, in Bott and Schlotterbeck's (2018) task it is probably easier to memorize a 1-to-1 pairing diagram than a diagram with an extra object (because the 1-to-1 pairing diagram without an extra object is less detailed). This results in an unfair comparison between the two diagrams as the working memory loads then differ between the conditions. The dual-task experiment conducted in the current study, in contrast, used a separate independent working memory task in the form of a digit span task, in which participants need to memorize a sequence of digits, independently from
the linguistic task (testing adults’ spreading errors). Each condition in the linguistic task is thus affected by a working memory load to the same degree, making it possible to directly compare the effects of this WM load on each condition. This suggests that Bott and Schlotterbeck (2018) are not in a position to draw the conclusion that spreading errors are induced by a limited working memory capacity, since their task did probably not limit the working memory capacity of the participants to a similar degree between the different conditions. They only made a regular truth-value judgment task harder. Their conclusion should therefore be that a more difficult truth-value judgement task, requiring the memorization of abstract set diagrams, seems to induce spreading errors in adults. The question then arises as to whether this is caused by the fact that pictures with an extra object might be harder to memorize than symmetric pictures without an extra object.

A second difference between the two studies concerns the form of the test sentences used. As previously mentioned, Bott and Schlotterbeck (2018) performed their experiment in German and used sentences like (14).

(14) Für jeden Schüler gilt: ihn lobte genau ein Lehrer voller Wohlwollen.
    For each pupil holds: him praised exactly one teacher full goodwill.
    Each pupil is such that exactly one teacher praised him full of goodwill.

These sentences are much more complex than the simple transitive sentences that are typically used in spreading research such as sentence (15) used in the present study.

(15) Elke jongen wast een boot.
    Each boy is washing a boat.

First, the sentence in (14) contains a subordinate clause and hence is syntactically more complex than (15). Second, in sentence (14) the universal quantifier jeden ‘each’ is part of the universally quantified expression jeden Schüler ‘each pupil’ in the matrix clause, which binds the fronted object pronoun ihn ‘him’ in the subordinate clause, whereas the universally quantified expression elke jongen ‘each boy’ in sentence (15) is not involved in any binding relation. Third, while the sentence in (14) contains the complex quantifier genau ein ‘exactly one’, sentence (15) contains the indefinite article een ‘a’. Finally, sentence (14) has an added optional adverbal phrase voller Wohlwollen ‘full of goodwill’, that adds to the complexity of the sentence.

Due to the differences in complexity, the sentences of Bott and Schlotterbeck (2018) are probably much harder to process than simple transitive sentences such as (15). Since Bott and Schlotterbeck (2018) used sentences like (14) it is not possible to rule out the fact that their sentences were simply too complex to be properly
processed. Note however that Bott and Schlotterbeck (2018) did perform an ordinary picture selection task (Experiment 1), without the cognitively demanding memorization of the diagrams in which they found only 2.3% spreading errors. In this picture selection task, participants neither had time pressure nor had to keep the diagram in memory and they could think about the sentence for as long as they needed (while at the same time looking at the picture). This was not the case during the cognitively demanding truth-value judgment task (Experiment 2: Bott & Schlotterbeck, 2018) in which segments of the sentence were displayed using a moving window presentation, since they had to memorize the diagram while processing the sentence. In short, this means that we cannot be certain that the errors made by the adults in Bott and Schlotterbeck’s (2018) experiment can be attributed to either spreading or task difficulty due to memorizing the pictures, or the complexity of the sentences.

![Diagram A](image1.png)

**Figure 10.**

- **a)** Original diagram used by Bott and Schlotterbeck (2018).
- **b)** Diagram in the same order as the arguments in the sentence.

A third issue related to the test materials of Bott and Schlotterbeck (2018) that does not play a role in the current study concerns the order of arguments in the sentences and the abstract set diagrams. The order of the sets in the diagrams from left to right was opposite to the linear order of the arguments in the sentence. For example, after memorizing a set diagram with teachers at the left and pupils at the right (Figure 10a), participants had to judge sentence (14), which first mentions pupils and then mentions teachers. If participants failed to notice this order difference, sentence (14) could be rejected because they may have incorrectly memorized the first set to represent the pupils and hence believe that there was an extra pupil (Figure 10b). Thus, rejections might indicate spreading, but could also be caused by incorrect diagram memorization. Please note that this order issue is discussed and tested in the third experiment presented by Bott and Schlotterbeck (2018). Crucially, however, in this experiment they did not test the classical spreading condition, so it is not possible to rule out the order issue as the cause of the spreading errors by the adults.
The previous paragraphs presented several differences between the de Koster et al. study (2018) and the study by Bott and Schlotterbeck (2018). The issues raised regarding the different dual-task procedure, the difference in complexity of the test sentences and Bott and Schlotterbeck's (2018) argument order issue, all seem to cast doubt on the way Bott and Schlotterbeck (2018) interpret their results.

Whereas the experiment of Bott and Schlotterbeck (2018) may have been too complex, one might wonder whether the dual-task experiment used in the current study may have been too simple for adults and not sensitive enough to load participants’ working memory. It might have been the case that we did not find spreading in adults for that reason. However, the results of conditions The-Extra Object and The-1-to-1 Pairing seem to indicate otherwise. Both conditions The-Extra Object and The-1-to-1 pairing showed acceptance rates higher than what was previously found by de Koster et al. (2018) using similar materials but without working memory load (27% vs. 73% and 25% vs. 73%, respectively, see Figure 8), pointing to an effect of working memory.

As mentioned in Section 4.5, condition The-1-to-1 Pairing is similar to condition The-Distributive examined in Chapter 5. Condition The-Distributive tested the implicature account for explaining adult distributivity interpretations. In short this account predicts that sentences with plural definites such as *the* are rejected by adults when they are combined with distributive pictures, due to the calculation of an implicature. Adult participants reason about an alternative sentence with the quantifier *each* that is a better fit to describe the distributive picture. They thereby infer that if the distributive interpretation was intended, the sentence should have contained *each*. This results in a collective preference for sentences with plural definites. The calculation of an implicature is assumed to require cognitive resources such as working memory. It is therefore predicted that condition The-Distributive, and thus The-1-to-1 Pairing, should be affected by working memory load. This is exactly what we found. Similar to the findings in Chapter 5, we found that the acceptance rate of condition The-1-to-1 Pairing under working memory load (73%) was much higher than the acceptance rate with no working memory load (25%, previously found by de Koster et al. (2018)).

The results for condition The-Extra Object are a bit harder to interpret, because a rejection can have different reasons, as explained in Section 4.5. A rejection in this condition could either indicate a spreading error, based on the extra object, or a collective preference. Condition The-Extra Object showed acceptance rates of 72% and 73% (in the low and high working memory load conditions, respectively) and these results perfectly mimic the results of conditions The-1-to-1 Pairing (from the current study) and The-Distributive (from the study presented in Chapter 5).
This makes us believe that the rejections in condition The-Extra Object are due to participants preferring a collective interpretation for sentences with plural definites and not due a focus on the extra objects. In short, the effects of WM load on conditions The-1-to-1 Pairing and The-Extra Object serve as evidence that our dual task was sensitive enough to load participants’ working memory.

7 Conclusion

The current chapter further examined the processing account of Musolino (2009) by investigating the weak-strong account (Geurts, 2003) on which it is based. Musolino’s processing account explains children’s non-adult-like distributivity interpretations by linking them to children’s spreading errors. This account is an extension of Geurts’ (2003) weak-strong account and offers a unified explanation of both children’s spreading errors and their non-adult-like distributivity interpretations. The weak-strong account connects children’s spreading errors to their limited cognitive resources. Previous work examining this account seems to be contradicting, with results pointing in different directions. De Koster et al. (2018) found no relationship between spreading and working memory, contrary to Bott and Schlotterbeck (2018). Bott and Schlotterbeck (2018) did find a relationship with cognitive resources. They were able to induce spreading errors in adults in a cognitively demanding version of a truth-value judgment task, which is in line with the predictions of the resource-based account of spreading.

To shed light on these contradicting results, the current study took a next step and conducted a dual-task experiment using the materials of de Koster et al. (2018). Our results did not match the results of Bott and Schlotterbeck (2018). Unlike Bott and Schlotterbeck (2018), we did not find spreading errors by adults when their working memory was loaded. These findings, together with several design issues in the Bott and Schlotterbeck study (2018), strengthen our belief that the results of the current study together with the results of de Koster et al. (2018) provide evidence against Geurts’ (2003) resource-based account of spreading and thus also against the extended processing account by Musolino (2009). The proposed relationship between children’s spreading errors and non-adult-like distributivity interpretations thus seems to be highly unlikely.