Concepts and Plural Predication

The Effects of Conceptual Knowledge on the Interpretation of Reciprocal and Conjunctive Plural Constructions

Published by

LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111

e-mail: lot@uu.nl
http://www.lotschool.nl

Cover photograph: Grand Prismatic Spring, Yellowstone National Park.
Taken by Eva B. Poortman, 2016.

ISBN: 978-94-6093-225-0
NUR: 616

Concepts and Plural Predication

The Effects of Conceptual Knowledge on the Interpretation of Reciprocal and Conjunctive Plural Constructions

Concepten en Predicatie over Meervouden

De Effecten van Conceptuele Kennis op de Interpretatie van Nevengeschikte Meervoudzinnen en Meervoudzinnen met Wederkerige Voornaamwoorden

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 3 februari 2017
des middags te 2.30 uur

door

Eva Barbera Poortman

geboren op 24 oktober 1986
te Maastricht

Promotoren: Prof. dr. Y. S. Winter
Prof. dr. M. B. H. Everaert

The research reported in this dissertation was supported by VICI grant number 277-80-002 by the Netherlands Organisation for Scientific Research (NWO).

Contents

Acknowledgements

Writing a book is something else! I did it, I'm proud, and I'm utterly grateful to the people who helped me along the way (and I think five years counts as quite a *long* way). So here goes.

The ones who were essential are first of all my promotors, Yoad Winter and Martin Everaert. Yoad, my daily supervisor, thank you for guiding me through this tough process, sharing your knowledge and your enthusiasm. Thank you for being available at odd hours and always willing to help out. I learned a lot from you and you made me feel supported, both as a researcher and as a person. Thank you for all the hours you put into bureaucracy on my behalf. Martin, my copromotor, thank you for being calm, for being down to earth and informed. Although you were mostly in the background during my PhD, it was essential to have you around, especially near the end of the project. Thank you for your support.

The third person who has played a major role in my PhD project is Liina Pylkkänen, who supervised me during my time at New York University. Liina, I still can't believe how much I've learned from you in the relatively short time of five months. It was great collaborating with someone as planned and organized as myself, and to experience how two organized minds work together and are able to complete a great project (*these girls are Dutch and Finnish*). Thank you for welcoming me in the NYU community and thank you for teaching me. I will never forget this once in a lifetime experience, nor the city I fell in love with.

For that experience and many more I am also grateful to NWO, UiL OTS, NYU, Maaike Schoorlemmer, Daphne van der Werf, Yvonne van Adrichem and Martien Camphuis. I gratefully acknowledge NWO VICI grant number 277-80-002 (awarded to Yoad Winter) and National Science Foundation Grant BCS-1221723 and grant G1001 from the NYUAD Institute, New York University Abu Dhabi (both awarded to Liina Pylkkänen). These institutions, people and grants made it possible for me to thrive in a great, intellectually challenging environment and to travel to places I never thought I'd get to see - both literally and figuratively speaking. Those experiences will stay with me forever. Thanks also to the friendly janitors and cleaning staff of Trans 10. I've enjoyed working in that old maze of a building.

A big “thank you” goes to my reading committee: James Hampton, Louise McNally, Lotte Hogeweg, Henriette de Swart and Frank Wijnen. Special thanks to those who have taken the time to comment on earlier versions of my work: Louise, James, Lotte and Joost Zwarts. Your feedback has been extremely valuable.

Doing a PhD is not possible without peers, without partners in crime, those who face the same troubles and difficulties as you do. Firstly, there are my paranimphs. Brigitta Keij, you are one of the few people whose company I always enjoy, no matter the mood I'm in. I still have the note you left for me on my very first day of work in 0.52, which you ended with "Brigitta (the messy desk behind you)". Some things never change! ;-) Thank you for listening, thank you for the nice talks, your good advice and your endless knowledge about whatever. Thank you for being my friend. Rob Zwitserlood, my other paranimph, shared an office with us for a big part of my PhD. Those were good times! Rob, thank you for your great company, lovely dinner parties (thanks also to Margo!), and wise words.

Other colleagues whose company I've enjoyed a lot over the years are Hanna de Vries (thank you for understanding my complaints and for being a great partner in organizing a conference), Assaf Toledo (thank you for keeping in touch), Marijn Struiksma (thank you for all your help and for listening during stressful moments), Marjolein van Egmond (I remember meeting you 7 years ago!), Anne van Leeuwen (you still need to come over for coffee), Heidi Klockmann (thank you for silliness), Anne-France Pinget (thank you for being someone to look up to), Yipu Wei (thank you for your thoughtful gifts), Sophia Katrenko, Choonkyu Lee, Sandrien van Ommen, Maartje Schulpen, Stavroula Alexandropoulou, Carien Westenbroek, Imke Kruitwagen, Robert Grimm, Anna Sara Romøren, Anja Goldschmidt, Jolien Scholten, Desiree Capel, Bjorn ‘t Hart, Mirjam Hachem, Liquan Liu, Liv Persson, Nadya Goldberg, Marko Hladnik, Lizet van Ewijk, Zenghui Liu, Loes Koring, Carolien van den Hazelkamp, Lotte Hendriks, Dominique Blok, Louise Nell, Ileana Grama, Suzanne Klein and Marieke Schouwstra.

Across the pond, in New York City, I had the pleasure of working with some great people, too. Thank you for your help and for making my time at NYU that much more pleasant: Jeff Walker, Miriam Lauter, Paul del Prato, Linmin Zhang, Phoebe Gaston, Masha Westerlund, Lucas Champollion and Chris Barker. And to Seohee Han, for being my friend in a strange, new place.

Thank you to some lovely people in the field that I met over the years: Galit Weidman Sassoon, Na’ama Friedmann, Heather Burnett, Lyn Tieu, Jacopo Romoli, Stephen Crain, Brian Buccola, Natalia Bekemeier, Polina Berezovskaya and Lena Karvovskaya. And thank you to Eddy Ruys and Riny Huybregts, for getting me excited about linguistics many, many years ago.

Finally, thank you to those who are often forgotten but are invaluable to experimental work: all research participants at Utrecht University and all brave MEG participants at

New York University. In addition, I am grateful to Ruth Noy Shapira, who illustrated all images for Experiment 3 and the experiment in Appendix C, as well as Figure 1 (Chapter 1), Figure 2 through 5 (Chapter 2), Figure 2 and 5 (Chapter 3) and Figure 2 and 3 (Chapter 5) in this dissertation.

These final sections are for my loved ones. Papa, mama and Felix, thank you for your endless support. During the past years I've realized that having a family, let alone feeling loved and supported by your family, is not something one should take for granted. So I choose not to. I am grateful for all the opportunities you've given me, for familiewap, for your trust in me, for your company, for *vlaoi*, and for always being there for me. Mama, thank you for late night phone calls, for moederdochterdagen (drie.. twee.. un), for your endless stream of good advice. Thank you for taking on the daunting task of reading my entire dissertation (!) and checking my English. Thank you for being extraordinary, and for making me want to be extraordinary, too.

Gerben, you must have had the hardest task in this process the past couple of years. I am positive that I could not have finished this book without you and am so lucky to have you. Thank you for comforting me when I couldn't see the end of this thing (which happened on multiple occasions), for pushing me to go on when I was sure I was going to quit (which usually followed), and for discussing my work with me when I needed it (still can't believe you read five versions of my Introduction chapter!). I'll never forget the amazing month-long roadtrip through the US that we made after I finished my manuscript (*there's a bison in the basin!*). You are the best partner I could hope for, you are my best friend. Dankjewel, lievie.

Chapter 1
Introduction

The focus of formal semantic studies of sentence meaning is on compositional operations of meaning, i.e. how meanings of separate elements within a sentence combine, and how the syntactic structure of the sentence affects that process. To study these operations, formal semanticists use insights from logic to translate a natural language sentence into a logical proposition, from which its truth conditions can be formulated. The common assumption is that only the function words in a sentence affect the logical proposition that it expresses. Function words form a closed class of words, including quantifiers (e.g. *every*), copulas (e.g. *to be*) and conjunctions (e.g. *and*), among others. Such words are distinguished from content words – verbs, adjectives, nouns – which form open classes to which new words can be added. The specific meanings of content words are assumed not to affect the derivation of the logical propositions derived in formal semantics. For example, given a sentence like *Sam is a man*, replacing the noun *man* with another noun such as *woman*, or even *dog*, obviously changes the logical proposition, but it does not have any effect on the way that that proposition is derived.

In this dissertation I argue that such a simple division of labor between function words and content words does not always hold, focusing on conjunctive and reciprocal plural constructions. The central claim throughout the work is that in addition to function words, the concepts expressed by content words can also affect the derivation of the logical proposition that such sentences express, hence their truth conditions. These effects are studied experimentally by measuring correlations between so-called *typicality effects* for verb concepts on the one hand and the attested logical propositions of sentences containing those concepts on the other hand.

1.1 Deriving logical propositions

Let us illustrate the process of generating a logical proposition in more detail, using sentence (1) below.

(1) John is sitting and reading

Sentence (1) contains the content words *John, sitting* and *reading*, and the function words *is* and *and*. The subject *John* refers to an individual, specified by the context of utterance. The verbs *sitting* and *reading* are predicates that specify the sets of entities that are sitting/reading in the relevant event. The copula *is* links the subject to the conjoined predicates: it partakes in the specification of tense and supports application of the complex predicate *sitting and reading* to its argument. Conjunction of predicates is analyzed as the intersection of the two sets denoted by the two verbs (e.g. Gazdar, 1980; Partee & Rooth, 1983; Keenan & Faltz, 1985). In (1), *and* intersects the set of sitting individuals and the set of reading individuals. Based on the meaning of the function words *is* and *and*, and the structure of the sentence, we create the following logical proposition: **john** ∈ (**sit** ∩ **read**). As a result, sentence (1) is modelled as true if John is in the intersection of the set of sitting individuals and the set of reading individuals, i.e. if John is sitting and John is reading, and false otherwise. Once we have formulated these truth conditions, we can test them against the actual world. If we know who John is and whether he is in the two relevant sets, we can determine whether (1) is true or false.

In the type of analysis just described, formal sentence meaning is derived independent of the exact meaning of content words. To illustrate: if we were to replace *John* in (1) by *Mary*, or *reading* by *singing*, the procedure deriving a logical proposition – be it **john** ∈ (**sit** ∩ **read**) or **mary** ∈ (**sit** ∩ **sing**) – would remain the same. For this reason formal semanticists often ignore many aspects of content word meanings.

1.2 Problems in the domain of plurals

From the discussion above we might tentatively conclude that content words do not take part in the analysis of formal sentence meaning. But now let us consider a plural sentence as in (2).

(2) The men are sitting and reading

Sentence (2) is quite similar to sentence (1), with the only difference being the fact that it asserts something about a plural (*the men*) instead of a singular (*John*). Plural sentences like (2) involve quantification over singular entities: to be true, the sentence surely requires that at least some individual men sit and at least some individual men read. However, specifying the truth conditions of sentence (2) more precisely is not an easy task. Perhaps (2) is about a small group of only two men, in which case we might take (2) to be true if both men are sitting and both men are reading, and false otherwise. This seems simple enough. But how about if only one of the men is sitting and the other man is reading? Is sentence (2) really false in such contexts? Furthermore, what happens when we are not talking about two, but four men? How many of those men should be sitting and how many should be reading in order for the sentence to be true?

And how about with six, ten or even a hundred men? The quantificational process in plural sentences like (2) is somewhat difficult to determine, and varies between speakers and situations, which makes it hard to determine exact truth conditions.

This dissertation presents experimental findings that quantitatively measure this kind of indeterminacy. As we will see, speakers show variation in their truth-value judgments on many plural sentences, including (2). For example, in one of the reported experiments, speakers were asked to make a truth-value judgment about (2) in a situation with four men in which only two are sitting and the other two are reading (Figure 1). In that situation, roughly 65 percent of the speakers considered (2) true, whereas the other 35 percent of the speakers considered it false. Such findings support the intuition that the quantificational process in (2) lacks an objective and precise description, but rather leaves room for some vagueness and variation between speakers. The aim of this dissertation is to clarify which parameters affect the level of indeterminacy that such sentences involve. To do that, it studies the interpretations of a variety of plural sentences, as well as potential factors affecting those interpretations. I will show that the choice of content words in plural sentences affects the truth conditions that speakers are likely to assign to them. This leads to a more general theory of the factors that govern the interpretation of plurals, and the way lexical meaning affects their formal semantics.

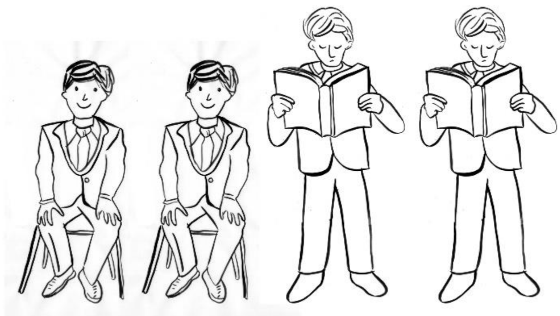

Figure 1: Situation with two men sitting and two men reading.

Previous works on vagueness with plural sentences have mostly concentrated on simple plural sentences like *The men are sitting*. In such a sentence, vagueness arises with respect to how many parts of the subject are affected by the predicate. If we assume that we are referring to four men, then *The men are sitting* is probably only true in case each of the four men is sitting. But if we are referring to one hundred men, it is quite natural to assert the sentence even when only 98 of them are sitting. Previous works have acknowledged this and specified a range of pragmatic factors that modulate the availability of such weaker, non-maximal interpretations (e.g. Westerstahl, 1984; Brisson, 1998; Lasersohn, 1999; Burnett, 2011; Schwarz, 2013; Križ, 2015). These factors include ‘domain restriction’ of the plural definite (e.g. *the men* refers only to the contextually relevant 98 men, not the total of 100) and 'pragmatic slack' (we allow a certain amount of imprecision, for example if the number of men is large enough, or if we perceive the men as a group). While vagueness is a generally accepted and well-studied characteristic of simple plural sentences, this dissertation focuses on vagueness in complex plural sentences that contain logical operators such as *and* and *each other*. As we shall see, the interpretation of such complex constructions is modulated by more factors than previous works have revealed. The aim of this dissertation is to specify these factors as precisely as possible using experimental methods, and to analyze the theoretical significance of these findings.

As mentioned above, speakers do not agree on the situations in which complex plural sentences like (2) are accepted. One possible account of this fact would be to suggest that such complex plural sentences are ambiguous. For instance, let us assume that *and* in (2) has both a strong, intersective meaning and a weaker meaning (which makes the sentence true in Figure 1). If speakers were choosing one of those meanings at random, a mixed interpretation pattern would be expected. If a speaker happens to interpret *and* as a strong intersection operator, they might reject (2) in the situation depicted in Figure 1. If, on the other hand, the speaker adopts the weaker meaning of *and*, they would accept (2). At first glance, one might think that this type of arbitrary ambiguity accounts for our experimental findings on sentence (2) in Figure 1. However, as this dissertation shows, arbitrary ambiguity is not enough to account for the interpretation of complex plural sentences. For example, let us reconsider sentence (2) (repeated below) and contrast it with sentence (3).

(2) The men are sitting and reading
(3) The men are sitting and standing

As already mentioned, (2) is accepted in a situation with two men sitting and two men reading (Figure 1) by 65% of speakers. If *and* is simply ambiguous, and its meaning is selected at random, then we would expect a similar mixed acceptability pattern for (3) in the logically parallel situation with two men sitting and two men standing. However, our results show that this is not the case: sentence (3) is accepted in such a situation by virtually all speakers (94%). Thus, the question arises why the weaker meaning is available to speakers for (3) much more often than for (2). A reasonable common sense explanation is that in contrast to (2), (3) is simply not imaginable under a strong, intersective interpretation of *and*, which would require each man to be sitting and standing simultaneously. Now, this kind of reasoning implies that the different meanings of content words (like *standing* and *reading*) have different effects on the available logical interpretations of complex plural sentences. That means that we require a non-random principle that resolves vagueness in a systematic way, taking into account differences between content words. Without such a principle, an ambiguity analysis of *and* cannot account for the differences between (2) and (3).

An influential principle in the domain of plurals that aims to take similar effects into account is the Strongest Meaning Hypothesis (SMH) by Dalrymple et al. (1998), which was adapted for plural conjunctions in Winter (2001a). The SMH selects the strongest meaning that is consistent with the linguistic and non-linguistic context of a sentence. In this way, the SMH allows logical meanings of sentences to differ not only due to different logical words or different sentence structures, but also due to a difference in intervening context. But how would it work? And what does 'context' mean?

Winter's (2001a) adaptation of the SMH deals specifically with plural sentences with predicate conjunction. Winter claims that such sentences sometimes receive a strong meaning and sometimes receive a weak meaning. For example, he argues that *The ducks are swimming and quacking* receives a strong meaning such that all ducks are swimming and all ducks are quacking, while the minimally different *The ducks are swimming and flying* receives a weak meaning such that some of the ducks are swimming and some of the ducks are flying. Since the logical words and structure of these sentences are identical, the specific contextual factor that modulates the different meanings has to be the meaning of the conjoined verbs *swimming and quacking* vs *swimming and flying*. Hence, whether the logical meaning of a sentence is strong or weak depends on the meaning of content words. According to Winter (2001a), sentences receive a strong meaning – where *and* means set intersection – unless such a meaning is strictly ruled out by the meanings of the conjoined predicates. If interpreting *and* as set intersection results in a sentence meaning that we cannot reasonably understand (e.g. in *The ducks are swimming and flying*), then the meaning of *and* is weakened. Otherwise, the meaning of *and* is set intersection (e.g. *The ducks are swimming and quacking*). Despite the fact that a formal semantic approach works for many types of sentences, with plural sentences 'logical' meaning is clearly affected by subtle 'non-logical' aspects of verb meaning. This seems to be on the right track, since we see something similar with the minimal pair (2) and (3). If we apply Winter's proposal to (3), then a strong, intersective meaning should be ruled out by the meanings of *sitting* and *standing*, hence (3) is expected to receive a weaker interpretation. However, if we apply the same approach to the analysis of (2), it is still not clear why some speakers allow a weak meaning. The meanings of the predicates *sitting* and *reading* surely do not exclude a strong meaning, i.e. we can reasonably imagine a situation in which four individual men are both sitting and reading simultaneously. So why do some speakers accept (2) in Figure 1? And, more generally, why do we find discrepancies among speakers? A reformulation of the SMH is necessary if we want to answer those questions.

1.3 Experimentally bridging the gap between sentence meaning and content word meaning

The experimental work in this dissertation reveals that even when content words in a sentence allow us to reasonably understand a logical operator in a plural sentence (e.g. *and*, *each other*) in a logically strong way, a percentage of speakers also accepts it in situations that are weaker. This goes against the predictions of the Strongest Meaning Hypothesis (Dalrymple et al., 1998; Winter, 2001a). This section outlines the argumentation as to why I think that is the case. To the aforementioned contrast between (2) and (3), let us add the additional example in (4) below.

(2) The men are sitting and reading
(3) The men are sitting and standing
(4) The men are sitting and cooking

As mentioned, (3) is virtually always accepted in a weak situation where each man is involved in only one activity. Sentences (2) and (4) show a more mixed pattern: quite a lot of speakers accept (2) and (4) in a weak situation, while others reject it. Crucially, a quantitative difference between (2) and (4) that my experiments reveal, hints as to why this is the case. Sentence (2) is accepted in a weak situation roughly 65% of the time, while (4) is accepted 85% of the time. At first glance, this is surprising. Relying on common sense, we can see why (3) behaves the way that it does – i.e. (3) does not make sense under a strong interpretation – but it is harder to see why (2) should behave differently from (4). Our world knowledge about both the verb combination *sitting and reading* and *sitting and cooking* allows situations in which each man is involved in both activities: it is, in principle, possible for a man to sit and cook simultaneously, just as it is possible for him to sit and read simultaneously. Despite this apparent similarity, the acceptability patterns of such pairs of sentences differ. This indicates that the knowledge that we need in order to explain the difference between (2) and (4) is more than about mere 'possibility' or 'impossibility' of situations.

Some previous works on complex plural sentences have also pointed out that interpretations of these sentences are quite variable (Philip, 2000; Winter, 2001a). This dissertation studies the source of that variability systematically - especially the sensitivity to the choice of content words – using a variety of experimental techniques. In this particular domain, experimentation has a clear advantage over informal collection of speaker judgments if one wishes to make reliable general claims about the phenomenon. As it turns out, without systematic experimentation, some of the data that I report could not be obtained. Firstly, it would be difficult to see that a plural sentence is not strictly true or false in a given situation (cf. (2) in Figure 1). Secondly, it would be virtually impossible to establish that sentences with seemingly similar content words have a different distribution of logical interpretations (e.g. (2) vs. (4)).

Such subtle aspects of the meaning of plural sentences can be established reliably only once such sentences are systematically tested in an experimental setting.

Once we know which small differences in sentence interpretation are systematically manifested in experiments, we can start addressing the question of *why* they show up. What differences between sentences like (2) and (4) indicate is that there must be very subtle aspects of the meanings of content words like *reading* and *cooking* that affect logical sentence meaning. In order to study them, we turn to insights on concepts from the cognitive psychology literature.

Since early behavioral studies of the 1970s, cognitive psychologists have extensively studied the *mental concepts* that content words refer to. Concepts allow us to categorize objects in the world. If we come across a bird, we are able to recognize it as such because we have a mental concept of 'bird'. Crucially, besides the information that allows us to categorize an object as a bird or not, concepts contain more detailed information. Within a category of instances that instantiate a concept, our minds appear to distinguish instances with respect to how 'good' a member they are. For example, even though both a robin and a penguin are categorized as birds, speakers consistently judge a robin as a better (or more typical) bird than a penguin. These are referred to as 'typicality effects': the psychological phenomenon that humans are able to rank different instances of a concept with respect to how representative those instances are for that concept (e.g. Rosch, 1973; Smith, Shoben & Rips, 1974; Rosch & Mervis, 1975). Typicality effects indirectly tell us something about the way that concepts are represented in a human mind. Apparently not all category members are represented equally.

Typicality effects have mostly been studied for concepts that can be expressed by nouns, but this dissertation shows that such effects also exist for concepts expressed by verbs. While nouns categorize entities (e.g. robin, penguin), verbs categorize situations (e.g. John is sitting, John knows Bill and George). Next, I make the non-self-evident assumption that typicality effects in categorization systematically affect the truth-conditions of plural sentences. With sentences like (2), (3) and (4), I show experimentally that a speaker's interpretation is affected by the typicality of the possible situations with respect to the relevant verb concepts. For example, the conjunction *sit and read* might not be different from the conjunction *sit and cook* in terms of the situations that are possible. However, the two conjunctions are different in terms of whether those situations are *typical* for the two concepts. For the combination *sit* and *read*, a situation in which John and Bill are both sitting and reading is just as typical as a situation in which only John is sitting and only Bill is reading. By contrast, for the combination *sit* and *cook*, experimental measures reveal that for many speakers a situation in which only John is sitting and only Bill is cooking is more typical than a situation in which John and Bill are both sitting and cooking simultaneously. I propose that it is these conceptual differences in the minds of speakers that explain the differences in interpretation of sentences like (2), (3) and (4) – both between sentences and between speakers.

The main research question throughout this dissertation is "*What mechanism accounts for the interaction between the logical interpretation of complex plural sentences and the non-logical meanings of content words?*". In answer to that, I propose a new hypothesis, called the Maximal Typicality Hypothesis (MTH), about the role of conceptual structure in the interpretation of complex plural sentences. The MTH encompasses fine-grained conceptual information on content words in terms of their typicality effects. For each complex plural sentence, the MTH selects one *core situation* based on that conceptual information: the maximal situation that is also maximally typical for the verb concept(s) in the sentence. In other words, the core situation is an optimal balance between what is maximal for the sentence and what is typical for the concept(s) in that sentence. For example, the maximal situation for sentence (4) is a situation in which each man is sitting and each man is cooking. However, since my results reveal that simultaneously sitting and cooking is atypical for the combination of concepts *sit* and *cook*, this cannot be the core situation for the sentence. Instead, the MTH selects a weak situation (where each man is involved in only one activity) as the core situation for (4), since this is the maximal situation that respects the concepts' typicality structure. Adding more activities to this situation would result in a situation where at least one individual is both sitting and cooking, which is atypical for the verb concepts. The MTH predicts that a sentence is highly acceptable in its core situation. Furthermore, the MTH also makes predictions about the acceptability of sentences in those situations that contain the core situation (i.e. those that contain more activities among the agents) and those situations that are contained within the core situation (i.e. those that contain fewer activities among the agents). The MTH is evaluated experimentally for conjunctive constructions such as *The men are sitting and cooking* and reciprocal plural constructions such as *John, Bill and George are pinching each other*, by measuring correlations between interpretation and typicality effects for verb concepts.

1.4 Structure of the dissertation

The dissertation consists of roughly two parts. The first part reports on experimental findings and theoretical implications concerning different types of complex plural sentences (Chapters 2, 3 and 4). Besides plural sentences with predicate conjunction as in (2)-(4), I study plural reciprocal sentences, which give rise to similar patterns. The second part of the dissertation is more conceptual and contains further issues and topics for further research (Chapters 5 and 6).

Chapter 2 deals with the challenge of reciprocal sentences: complex plural sentences with the logical operator *each other* such as *John, Bill and George are pinching each* *other*. The MTH is introduced as the mechanism that accounts for the interaction between logical plural sentence meaning and conceptual aspects of content word meaning. I address the problems of experimentally testing the MTH, i.e. testing typicality effects with verb concepts like *pinch*. ***Experiment 1*** experimentally tests the MTH. The experiment contains two parts: a) a typicality test that measures typicality preferences with different verb concepts and b) a truth-value judgment test with reciprocal sentences containing those verb concepts. The first part demonstrates that the typical situations vary between verb concepts, and the second part shows a significant effect of this variation on reciprocal interpretation: the stronger the concept's bias for a situation with one patient per agent, the weaker the interpretation of reciprocal sentences containing it – supporting the MTH.

In **Chapter 3**, I study another challenge in the domain of plurals: complex plural sentences with predicate conjunction such as *The men are sitting and cooking*. Based on findings in Chapter 2, the MTH is expected to also account for these sentences. Again, I argue that the MTH is the mechanism that accounts for the interaction between sentence meaning and word meaning. I address the problems of experimentally testing the MTH in this domain, i.e. testing typicality effects for verb concept combinations like *sitting and cooking*. With two experiments, the MTH is revealed as a viable principle for sentences with predicate conjunction just like it proved to be for reciprocals in Chapter 2. ***Experiment 2*** refutes a previous account, the Strongest Meaning Hypothesis, which suggests that such sentences are always interpreted intersectively, unless this is strictly excluded by properties of the predicates (Winter, 2001a). ***Experiment 3*** tests the MTH. Again, this requires two parts: a) a typicality test that measures preferences with different pairs of verb concepts and b) a truth-value judgment test with sentences containing those pairs. Part one demonstrates that the typical situations vary between different pairs of verb concepts, and part two shows a significant effect of that variation on sentence interpretation: the stronger the pair's bias for a situation in which each individual is involved in only one activity, the weaker the interpretation of plural sentences containing that pair in conjunction – again supporting the MTH.

Chapter 4 uses findings on predicate conjunction from Chapter 3 in a neurolinguistics study on a specific brain region that has been implicated in the composition of complex concepts: the Left Anterior Temporal Lobe (LATL). ***Experiment 4*** contrasts sentences (from Experiment 2) that are always interpreted intersectively (e.g. *The animals are big and gray*) with sentences that are never interpreted intersectively (e.g. *The animals are big and small*). Results reveal that the LATL is more active when speakers process intersective sentences, where two properties apply to each individual, than when they process non-intersective ones, where only one property applies to each individual. This allows us to more precisely characterize the computational contribution of the LATL in semantic processing, and it shows more generally that complex plural sentences show differences not only with behavioral measurements, but also during processing.

While previous chapters have dealt with the effects of lexical knowledge on sentence interpretation, **Chapter 5** offers a discussion on the effects of a different type of contextual parameter, namely spatial knowledge. The chapter contains preliminary experimental results that reveal the effects of spatial knowledge on the interpretation of reciprocal sentences, and speculates about similar effects on the interpretation of plural sentences with predicate conjunction.

In **Chapter 6,** I discuss how the results on lexical-conceptual effects from Experiments 1 through 4 relate to concept composition. The complex plural sentences in the center of this dissertation and modification constructions in the center of the literature on concept composition show similar types of meaning conflicts. This chapter explores the possibility of having a general mechanism of conflict resolution that captures both phenomena.

Chapter 7 concludes the dissertation.

Chapter 2
Reciprocity & typicality effects

Part of this chapter also appeared in Poortman, Struiksma, Kerem, Friedmann and Winter (under revision).

2.1 Introduction

This chapter studies the interpretation of reciprocal sentences. Simple reciprocal sentences, also called elementary reciprocal sentences (Langendoen, 1978), consist of three elements: 1) a reciprocal expression *r*, for example the English expressions *each other* or *one another*, 2) an NP denoting antecedent set *A* of cardinality two or more, for example *the boys* or *John and Bill*, and 3) a relation *R*, for example *know* or *pinch*. Elementary reciprocal sentences are of the form *ARr*. Examples are in (1)-(3).

(1) John and Bill know each other
(2) The boys know each other
(3) John, Bill and George know each other

Reciprocal sentences provide speakers with a concise way of describing complex situations. Sentence (3), for example, naturally describes a situation with six acquaintance relations: John knows Bill, Bill knows John, John knows George, George knows John, Bill knows George and George knows Bill, i.e. every man knows every other man. Such a paraphrase of all the separate statements entailed by the reciprocal sentence expresses an interpretation of that reciprocal sentence, e.g. a situation that supports a truthful use of the sentence. I will henceforth refer to any situation that demonstrates full mutuality between three agents (as in (3)) as an *S6* situation. Since Langendoen's early work, many studies of reciprocals in formal semantics have presented compelling evidence that despite their similar structure, different reciprocal sentences can describe different situations (e.g. Dalrymple et al., 1998; Beck, 2001; Sabato & Winter, 2012). These differences arise as soon as the antecedent set *A* contains more than two individuals. Consider for example sentence (4), which is only minimally different from sentence (3).

(4) John, Bill and George are biting each other

The logical structures of (3) and (4) are similar. Sentence (4) only differs from sentence (3) with respect to the relation *R* in the scope of the reciprocal expression and with respect to tense. Based on the interpretation of sentence (3), we might expect that (4) would describe a situation in which every man bites every other man (i.e. *S6*). However, this is not the case: sentence (4) most likely describes a situation with only three relations, where every man is biting only one other man. I will refer to such situations as *S3* situations.

The main question throughout this chapter is what explains such differences between reciprocal sentences. In order to answer that question, it studies the contextual-lexical effects on the logical semantics of simple reciprocal sentences in detail.

In an influential work, Dalrymple et al. (1998) propose a way to explain the different interpretations of reciprocal sentences such as (3) and (4). Using a principle that they call the Strongest Meaning Hypothesis (SMH), Dalrymple et al. propose that each occurrence of a reciprocal expression denotes one of six different logical operators. Each of these operators encodes a different formal mechanism for categorizing situations. For instance, the operator of *Strong Reciprocity* (SR) categorizes those situations where every member of the antecedent set is connected by the given relation to every other member. Thus, in sentences like (3) and (4) where the antecedent set consists of three members, applying SR results in selecting *S6* as the only possible situation where the sentence can be asserted truthfully.[1] The other five operators that Dalrymple et al. propose besides SR use different logical requirements to categorize reciprocal situations. The SMH is proposed as the principle that selects between Dalrymple et al.'s six logical operators. The selection is performed using assumptions about the linguistic and non-linguistic context in which the reciprocal expression is used. For each occurrence of a reciprocal expression in a given context, the SMH selects the operator that results in the *strongest sentential meaning* that is consistent with that context. Other operators are ruled out. For example, in sentence (3) Dalrymple et al. assume that context puts no restrictions on the number of possible acquaintances that any of the three persons may have.[2] As a result, the SMH selects the strongest sentence meaning in this context, where *each other* means SR: every man knows every other man. This analysis predicts that sentence (3) can only be true in an

[1] Formally, it should be noted that *S6* actually represents a class of situations where each agent acts on both other agents, and possibly itself. Because this formal detail is irrelevant for the current discussion, I sloppily refer to *S6* as one situation.

[2] For consistency with Dalrymple et al.'s terminology, I refer here to the world knowledge about predicates like *know* and *bite* as being part of the 'context'. Whether this term is adequate depends on one's conception of lexical meaning and context, but it is irrelevant for the current purposes.

S6 situation, which is in line with many speakers' intuitions. Sentence (4) is also successfully analyzed by the SMH, if we make the plausible assumption that each individual can only bite one other individual at a time. Taking this assumption on the meaning of *bite* to be part of the context of (4), the SMH cannot select SR for the reciprocal expression in (4). This would result in an *S6* situation, which would be inconsistent with the context. Consequently, a logically weaker operator is selected. This operator correctly analyzes sentence (4) as true in all *S3* situations, where each individual bites only one other individual (e.g. John bites Bill, Bill bites George and George bites John).

Thus, we see that based on plausible contextual assumptions, the SMH correctly accounts for the intuitive distinction between sentences (3) and (4). This account is based on the assumption that the verb *bite* restricts the number of possible simultaneous patients per agent, whereas the verb *know* does not. More generally, in all cases that Dalrymple et al. discuss, their analysis relies on the assumption that the context categorically restricts the number of patients that an agent may possibly have simultaneously. This assumption is needed in order for the SMH to allow reciprocal meanings that are weaker than Strong Reciprocity. In sentence (4), this assumption looks innocuous enough: it looks reasonable to assume that most ordinary contexts prohibit agents from biting more than one patient simultaneously. However, with many other sentences, this kind of assumption is more questionable. For instance, let us consider sentence (5).

(5) John, Bill and George are pinching each other

In terms of the situations that it supports, sentence (5) does not differ much from sentence (4): intuitively, both sentences seem equally true in *S3* situations. However, despite this similarity in the sentential interpretations of (4) and (5), the predicates *pinch* and *bite* are quite different in terms of the meaning restrictions that they induce on possible contexts. While in natural contexts, biting two people simultaneously seems impossible, pinching two patients simultaneously is physically possible. As a result, in normal contexts where people use both their hands, the SMH would select SR as the meaning of the reciprocal in (5). Consequently, the SMH expects (5) to be acceptable only in the *S6* situation, where every man pinches every other man. This prediction is problematic. Both experimental work by Kerem, Friedmann and Winter (2009) and the experimental work reported in this chapter indicate that the prediction is not borne out. Kerem, Friedmann and Winter (2009) show that when participants are asked to choose between *S3* and *S6* for a given reciprocal sentence such as (5), where both situations are realistically depicted with human agents, they often prefer *S3* over *S6*. In such contexts, the illustrated possibility of *S6* leads the SMH to predict that SR is the only reciprocal reading. Accordingly, the SMH expects the reciprocal sentence to be true in *S6* and false in *S3*. Thus, it is predicted that *S3* should never be preferred over

S6, contrary to fact. In our[3] work, we support this finding by showing that sentences like (5) are accepted in *S3* by most speakers in a truth-value judgment task. These results are a major challenge for Dalrymple et al.'s use of the SMH. We hypothesize that speakers take into account the implausibility of situations with two patients per agent (like *S6*) with predicates like *pinch*, even when they are evidently aware of their physical possibility.

Dalrymple et al., somewhat non-standardly, assume that sharp restrictions on verb meanings – e.g. whether a person can or cannot pinch more than one object at a time – are introduced as part of the context of utterance. In this chapter, I reject the assumption that only such sharp restrictions on meaning play a role in the interpretation of reciprocal sentences. Rather, I focus on a more general factor: the conceptual structure of the relational concept that the verb refers to. As an alternative to the SMH, a new selection principle is proposed that accounts for the interaction between reciprocal quantifiers and verb concepts. The proposed account predicts the truth-value of a reciprocal sentence in a given situation not based on sharp conceptual distinctions, as in Dalrymple et al.'s proposal, but in correlation to fuzzy ones: the typicality of the situation for verb concepts like *know*, *pinch* or *bite*. It is based on a principle called the *Maximal Typicality Hypothesis* (MTH), which is formulated based on initial work by Kerem, Friedmann and Winter (2009). For any reciprocal sentence, the MTH designates one situation as the so-called *core situation*: the situation that contains enough relations between the agents in the sentence to satisfy reciprocity, and thus make the sentence true. Crucially, what we define as "enough relations" critically depends on typicality information about the verb concept in the sentence. In contrast to the workings of the SMH, the core situation is not necessarily the only situation in which a sentence is acceptable. Instead, it functions as a boundary: a reciprocal sentence is predicted to be true in the core situation and all situations that properly contain the core situation. The MTH qualitatively explains the experimental results that are problematic for the SMH and other proposals about the meaning of reciprocals. Furthermore, the MTH makes novel and more precise quantitative predictions about the relations between reciprocity and typicality. The novel experimental work that I report supports the MTH as a viable principle governing reciprocity in natural language.

The chapter is structured as follows. Section 2.2 briefly reviews some of the research that was done in the area of reciprocals over the years. The work will be categorized roughly into three different types of research questions. The first question concerns which different logical operators the reciprocal expression can possibly denote. I discuss the most systematic works that survey different operators in order to demonstrate the variety of logical contributions that a reciprocal expression could in principle make to sentence meaning. Secondly, for completeness I briefly mention some

[3] While this chapter has one author, the work described in sections 2.3 and 2.4 was carried out together with Marijn Struiksma and Yoad Winter.

works dealing with questions on the syntax/semantics interface of reciprocals, i.e. how we derive different sentence meanings for reciprocal sentences with different syntactic structures. Finally, the section zooms in on the semantic/pragmatic questions that are in the focus of this dissertation. I will be mainly interested in which lexical-semantic or pragmatic principles affect the availability of certain reciprocal interpretations. Section 2.3 then introduces the new formulation of the Maximal Typicality Hypothesis (MTH). Section 2.4 describes new experimental work that systematically tests the predictions made by the MTH, and section 2.5 concludes this chapter.

2.2 Some background on reciprocals

This section briefly reviews some of the literature on reciprocals. A distinction has been made between descriptive, syntactic/semantic and semantic/pragmatic questions, although one should of course keep in mind that most of the works that will be discussed deal with a complicated mix of these questions. Having acknowledged that, I try to untangle them into the three categories in order to carve out the specific area that this chapter is concerned with.

2.2.1 What does each other mean?

The first question is descriptive and is discussed in order to demonstrate how diverse the area of reciprocity actually is, without assuming any specific theoretical account. Before one can study how different interpretations of reciprocal sentences come about, it is useful to get a sense of which sentence meanings can actually be recognized empirically and thus which logical contributions the reciprocal expression can make. This has turned out to be a more difficult task than one might envision, and is apparent through the different inventories of logical operators that were proposed for the reciprocal expression *each other* in the literature. To get a sense of that variability, I here discuss the inventories of operators that have been discussed in Langendoen (1978) and in Dalrymple et al. (1998), since these are descriptively the most diverse overviews in the literature.[4]

---

[4] One important theme in the literature is the precise way in which the different interpretations of reciprocals are derived from general assumptions about plurals. For some notable examples of works that study this topic see: Schwarzschild (1996), Sternefeld (1998), Beck (2001). I will touch upon some of these works in section 2.2.2 below. However, the current discussion is mainly geared by the motivation of selecting between those different interpretations in any particular sentence, and not by the precise way in which they are derived. For this reason the discussion focuses on Langendoen's and Dalrymple et al.'s works, which set the empirical stage for many further works on the syntax and semantics of reciprocals.

Langendoen (1978)

Langendoen (1978) was the first to come up with an elaborate inventory of reciprocal operators. He considers six, logically distinct, “not unlikely potential schematic representations of the truth conditions for elementary reciprocal sentences” (p. 178). This section describes the six different possibilities that Langendoen came up with, and presents an example for each one.

Firstly, **Strong Reciprocity (SR)** is probably the most straightforward operator that the reciprocal can denote. It says that every member of *A* is related directly to every other member of *A* by the relation *R* (possibly except for itself).

Strong Reciprocity (SR) (also called *each-the-other* by Fiengo & Lasnik, 1973) $(\forall x,y \in A)\ (x \neq y \rightarrow xRy)$

Consider again example (3). The reciprocal expression in (3) is assumed to be an operator of the type SR, since sentence (3) is interpreted such that every member of *A* (*John*, *Bill*, *George*) is directly related to every other member of *A* (possibly except itself) by the relation *know*. In other words, every man knows every other man: John knows Bill and George, Bill knows John and George, and George knows John and Bill.

Related to SR is **Partitioned Strong Reciprocity (PSR)**, which says that if *A* can be divided into subsets $A_1, \ldots, A_n$ such that each subset has cardinality two or more, then every member of a subset is related directly to every other member of that subset (except itself) by the relation *R*. In other words, PSR is simply SR holding over disjoint subsets that cover the whole set *A*.

Partitioned Strong Reciprocity (PSR) (also called *reciprocal reciprocity* (Fiengo & Lasnik, 1973) and *distinct subsets reciprocity* (Dougherty, 1974)) Let *A* be partitioned into subsets as follows: $A = A_1 \cup \ldots \cup A_n$ and $(\forall i, j, 1 \leq i, j \leq n)\ (i \neq j \rightarrow A_i \cap A_j = \emptyset)$ and $(\forall k, 1 \leq k \leq n)$ (card $A_k \leq 2$) $(\forall i, 1 \leq i \leq n)\ (\forall x, y \in A_i)\ (x \neq y \rightarrow xRy)$

Fiengo and Lasnik (1973), who first proposed this reciprocal reciprocity, as they called it, claim that it correctly captures the truth conditions of sentence (6) in a situation in which there are nine men in a room, divided into three groups of two and one trio.

(6) The men are hitting each other

Fiengo & Lasnik claim that (6) is a correct description of a situation in which within each of the four subsets of men (that make up the full antecedent set *A* of nine men), every man hits every other man.

Thirdly, **Symmetric Reciprocity (SymR)** says that every member of *A* has at least one 'partner' in *A* with whom the relation *R* holds in both directions.

Symmetric Reciprocity (SymR) (also called *unrestricted subsets relation* by Dougherty (1974) and *Conjunctive Ordering* by Kański (1987)) $(\forall x \in A)\ (\exists y \in A)\ (x \neq y \wedge xRy \wedge yRx)$

Such an operator has been ascribed to the reciprocal expression in a sentence like (7) by Dougherty (1974), which he claims can be true even if only heterosexual relations are involved.

(7) John, Bill, Tom, Jane and Mary had relations with each other

According to Dougherty, sentence (7) can be true if Mary had relations with Bill and Tom, but Bill and Tom did not have relations with each other.

The fourth operator by Langendoen is **Intermediate Reciprocity (IR)**. IR says that every member of *A* is directly or indirectly related to every other member of *A* (except itself) by a directed path described by the relation *R*.

Intermediate Reciprocity (IR) (*directed connectedness* of graph theory) $(\forall x,y \in A)\ (x \neq y \rightarrow [xRy \vee (\exists n > 0)(\exists z_1, \ldots, z_n \in A)(xRz_1 \wedge \ldots \wedge z_nRy)])$

An example is given by Dalrymple et al. (1998), and is repeated here in (8).

(8) The telephone poles are spaced five hundred feet from each other

Sentence (8) is assumed to be true when every member of *A* (pole 1, pole 2, ..) is directly or indirectly related to every other member of *A* (except itself) by the relation *to be spaced five hundred feet from*. In other words, the sentence is true in a situation in which each pole is five hundred feet from the nearest one(s), and false if two adjacent poles are separated by a distance other than five hundred feet (Dalrymple et al., 1998).

Next, **Partitioned Intermediate Reciprocity (PIR)** is similar to PSR in the sense that the sentence's truth conditions are formulated based on the relations within disjoint subsets that cover the set *A*. PIR says that if *A* can be divided into subsets $A_1, \ldots, A_n$ such that each subset has cardinality two or more, then every member of a subset is directly or indirectly related to every other member of that subset (except itself) by the relation *R*. In other words, PIR is simply IR holding over disjoint subsets that cover the whole set *A*. Neither Langendoen nor other works that discuss this proposed operator provide examples for PIR.

Partitioned Intermediate Reciprocity (PIR)
Let A be partitioned into subsets as for PSR above.
$(\forall i, 1 \leq i \leq n)\ (\forall x, y \in A_i)\ (x \neq y \rightarrow [xRy \vee (\exists n > 0)(\exists z_1, \ldots, z_n \in A)(xRz_1 \wedge \ldots \wedge z_nRy)])$

Finally, Langendoen's **Weak Reciprocity (WR)** is the one that he claims to be the weakest of the six. WR says that every member of *A* is related to at least one other member of *A* by the relation *R* as the first argument, and is related to at least one other member of *A* by the relation *R* as the second argument.

Weak Reciprocity (WR)
$(\forall x \in A)\ (\exists y,z \in A)\ (x \neq y \wedge x \neq z \wedge xRy \wedge zRx)$

As an example for WR, Langendoen gives sentence (9), in a context where *A* consists of five individuals: two of the individuals weigh 50 kg each, one weighs 60 kg, and the remaining two weigh 70 kg each.

(9) They are at least as heavy as one another

If the reciprocal expression in sentence (9) is indeed of the type WR, then the sentence is true in case every member of *A* (individual 1 .. individual 5) is related to at least one other member of *A* by the relation *to be as heavy as* as the first argument, and is related to at least one other member of *A* by that relation as the second argument. Reconstructing Langendoen's claim, this would mean that for every individual, there is at least one other individual who is the same weight or lighter, and there is at least one other individual who is the same weight or heavier. Langendoen (1978) himself acknowledges that sentences such as (9) are "felt to be bizarre independent of the truth or falsity of the assertions that they make (for example, even if the five individuals weigh exactly the same amount, so that it makes an indisputedly true assertion, the sentence still seems strange)" (p. 183). I will come back to this type of sentence later on in this section, where Dalrymple et al.'s reaction to Langendoen's WR is discussed.

Dalrymple et al. (1998)

Dalrymple et al. (1998) also provide an overview of logical operators for the reciprocal, by presenting the range of meanings that they have found empirically. They start their paper from the observation that "*each other* and *one another* vary in meaning according to the meaning of their scope and antecedent as well as the context in which they are uttered" (p.159). The paper then continues with a taxonomy of meanings that they claim are attested in natural language. They take two of the operators that Langendoen proposed as potential meanings to be genuine meanings, namely Strong Reciprocity and Intermediate Reciprocity. On top of that, they add one operator that was previously proposed by Kański (1987) and two operators that have not been proposed before –

resulting in a total of five[5] different operators. The latter three operators are described below.

Firstly, **One-way Weak Reciprocity (OWR)** says that every member of *A* participates in the relation *R* with at least one other member of *A* as the first argument.

One-way Weak Reciprocity (OWR) $(\forall x \in A)\ (\exists y \in A)\ (x \neq y \rightarrow xRy)$

An example for this type of operator that Dalrymple et al. provide is in (10).

(10) "The captain!", said the pirates, staring at each other in surprise

According to the authors, sentence (10) requires merely that each pirate stares at at least one other pirate. There are no requirements regarding the second argument, meaning that it is not necessary for each pirate to be stared at by another pirate – hence the *One-way* WR.[6]

Secondly, **Intermediate Alternative Reciprocity (IAR)** "requires that all pairs in *A* be connected directly or indirectly via the relation *R*, ignoring the direction of the arrows" (Dalrymple et al., 1998, p.173).

Intermediate Alternative Reciprocity (IAR) (*non-directed connectedness* of graph theory) $(\forall x,y \in A)\ (x \neq y \rightarrow$ for some sequence $z_0, \ldots, z_m \in A)(x = z_0 \wedge (Rz_0z_1 \vee Rz_1z_0) \wedge \ldots \wedge (Rz_{m-1}z_m \vee Rz_mz_{m-1}) \wedge z_m = y))$

An example given by Dalrymple et al. is in (11).

(11) The third-grade students gave each other measles

Sentence (11) is true if all pairs of students in *A* are connected directly or indirectly via the relation *to give measles to*. That means that each student gave a class member measles and/or got the disease from a class member.

Finally, **Inclusive Alternative Ordering (IAO)** is a very weak operator and was first proposed by Kański (1987). It requires merely that every member of *A* is related

[5] Dalrymple et al. (1998) first propose five different operators. However, based on the logical parameters of variation between those five operators, they eventually assume that there must be a sixth possible meaning for reciprocals that they did not attest: Strong Alternative Reciprocity (see p.186-188 in Dalrymple et al. (1998)).
[6] In addition, (10) seems to require that more than one pirate was stared at, as pointed out to me by J. Hampton (personal communication).

to at least one other member of A by the relation R as the first or as the second argument or both.

Inclusive Alternative Ordering (IAO) (Kański, 1987)
$(\forall x \in A)\ (\exists y \in A)\ (x \neq y \wedge (xRy \vee yRx)$

An example from Dalrymple et al. (1998) is in (12).

(12) The inmates slept on foot-wide wooden planks stacked atop each other

Dalrymple et al. claim that such a sentence is true in a situation where there are several stacks of planks. As long as each plank is either on top of or below another plank, the sentence is true, thus not requiring all planks to be connected.

Moreover, Dalrymple et al. claim that what Langendoen called Symmetric Reciprocity (SymR) is in fact empirically speaking equivalent to IAO. That means that the examples that Langendoen gives in favor of SymR (e.g. example (7) above) are also captured by IAO. According to Dalrymple et al., they are simply a special case of IAO, namely one with inherently symmetric relations. They use the same argument to claim that Langendoen's Weak Reciprocity (WR) is also captured by IAO. The examples given in the literature are again examples with inherently symmetric relations, such as the one in example (9). In sum, the examples given as proof for WR and SymR are best regarded as exemplifying IAO, they say, "since IAO is also exemplified by nonsymmetric cases whose meaning is definitely not any of the other definitions" (p.176).

Interim summary

This section described some of the logical operators for the reciprocal expression that the literature has exposed. Langendoen (1978) describes six different operators as potential candidates for the reciprocal, while Dalrymple et al. (1998) claim to have evidence for five. Between the two taxonomies, only two of the proposed operators overlap. The aim of section 2.2.1 was to demonstrate the diversity of different logical contributions to sentence meaning that the reciprocal expression *each other* could in principle make. This dissertation is not directly concerned with the question of which operators are the 'correct' ones, but with the remaining problem of what regulates them: can we find testable parameters that explain the variability that the literature has exposed?

2.2.2 How do we derive different sentence meanings systematically from syntactic structure?

Before moving to the main question of this chapter, namely what regulates different meanings of reciprocal sentences, the current subsection briefly overviews some of the syntactic approaches to reciprocal meaning. Besides the question of what reciprocal sentences can mean, an important question that has been asked is how those meanings are derived from the syntactic structure of the reciprocal sentences. The brief discussion below is by no means an exhaustive literature review in this area of research. Its main aim is to put the problems that drive the current work in the context of other outstanding questions about the syntax/semantics interface of reciprocals.

Heim, Lasnik & May (1991)

Heim, Lasnik and May claim that, semantically, reciprocals play two distinct roles (observed by Bennett, 1974). The first is distribution, i.e. they introduce universal quantification. The second is reciprocation, i.e. they require distinctness, such that only assignments of distinct individuals are considered as satisfying the relation to which the reciprocal applies (Heim, Lasnik & May, 1991, p. 67). We have already seen both these roles arise in the operators that were discussed in the previous subsection. Firstly, through the universal quantifier (□) in the formal statements and secondly, through the x □□y statement, indicating that relations between members of a set A are required to hold as long as they are not between a member and itself. Heim, Lasnik and May claim that these two roles come from the inherent semantics of the nonreciprocal uses of *each* and *other*: the term *each* is the distributor while *other* is the reciprocator. The meaning of reciprocals, in turn, arises from the compositional interaction of the meanings that their constituent parts, *each* and *other*, have in isolation.

On the syntactic side of their account, the authors start from the assumption that part of the reciprocal pronoun *each other* undergoes movement when it is mapped onto a Logical Form representation (e.g. Lebeaux, 1983; Chomsky 1986). Specifically, they assume that “a marker of distribution, the element *each*, is removed from its surface position and adjoined at LF to its “antecedent” phrase” (p.66). Thus, *each* moves to the position in which it distributes. Different meanings of reciprocal sentences arise due to the fact that *each* can adjoin at several different positions in the syntactic structure. Take for example sentence (13), which is a more complicated reciprocal sentence than we have seen so far, taken from Heim, Lasnik and May (1991).

(13) John and Mary told each other that they should leave

The authors claim that the reason that sentence (13) is multiply ambiguous, is that *each* can move to three different positions in its underlying structure: there are three ways to anaphorically bind it. It can be bound relative to the distribution index contributed by *each*, resulting in the “I” reading (each told the other “I should leave”), it can be bound relative to the index of *other*, resulting in the “you” reading (each told the other “You should leave”), and it can be bound relative to the range index of the antecedent subject NP, resulting in the “we” reading(s) (each told the other “We should leave”).

Sternefeld (1998) and Beck (2001)

In contrast to Heim, Lasnik and May (and many others such as Chomsky, 1973; Dougherty, 1974; Lebeaux, 1983), Sternefeld (1998) as well as Beck (2001) do not take the reciprocal itself to be responsible for distributivity. Instead, the quantification that Heim, Lasnik and May associate with *each* is claimed to not be derived from a process like *each*-movement, but instead analyzed as part of the semantics of pluralization in general – an insight going back to Langendoen (1978). These works propose that reciprocal sentences are a kind of relational plural sentences, as in (14), taken from Beck (2001).

(14) The children touch the horses

Relational plural sentences contain two group denoting expressions (e.g. *the children*, *the horses*) and a relation (e.g. *touch*). Sternefeld and Beck claim that reciprocal sentences are similar: “the difference comes from the anaphoric nature of the reciprocal: the same group is quantified over twice in the case of a reciprocal sentence” (Beck, 2001, p. 78). Relational plural sentences like (14) have an interpretation involving universal quantification over both groups in the sentence, e.g. all children touch all horses. This is the parallel to SR for reciprocal sentences in which the same group is quantified over twice. In addition, such sentences can also have cumulative interpretations which are weaker, e.g. each child touches one of the horses, and each horse is touched by one of the children. This weaker reading parallels weaker reciprocal meanings. These authors claim that since reciprocal sentences are simply a special kind of relational plural, the variability in their interpretation must be due to the same indeterminacy that we observe in relational plural sentences. In order to explain that variability, they therefore make use of the same mechanisms of plural predication that are independently needed for relational plurals: “pluralization operators, LF operators like QR, and addition of contextual information” (Beck, 2001).

2.2.3 Which principles affect the availability of certain sentence interpretations?

The previous sections have discussed different angles of research on reciprocals, namely research from a descriptive semantic angle and from a syntax/semantics interface angle. While the relevance of those angles is acknowledged, the focus of this chapter is of a more pragmatic nature. I am interested in which contextual principles play a role in the availability of certain interpretations. An influential paper dealing with this question regarding reciprocal sentences is one that was already mentioned before, namely Dalrymple et al.’s (1998) paper. Dalrymple et al. go further than providing an inventory of possible meanings for the reciprocal expression. They also propose a principle that selects one of those meanings for each occurrence of the reciprocal expression. Others who have dealt with similar questions are Sabato and Winter (2012) and Mari (2013). This section discusses these works in some detail, with a focus on Dalrymple et al.'s Strongest Meaning Hypothesis, since this proposal has served as a basis for subsequent work, including the current dissertation.

Dalrymple et al. (1998): Strongest Meaning Hypothesis

In section 2.2.1 Dalrymple et al.'s inventory of six reciprocal operators was discussed. Making such a list, however, was not the main goal of their work. The authors aimed first and foremost to account for the variation in reciprocal sentence meanings that they attested and propose that context needs to be taken into account – which, in their proposal, includes the verb in the reciprocal's scope as well as further linguistic and extra-linguistic information. Dalrymple et al. were the first to incorporate an explicit notion of context within logical meaning of reciprocals in this way. They propose a principle that selects one of their logical operators for each occurrence of the reciprocal, based on its context: The Strongest Meaning Hypothesis (SMH). Consider again the sentences (3) and (4), repeated below as (15) and (16).

(15) John, Bill and George know each other
(16) John, Bill and George are biting each other

As was discussed in the introduction, sentences (15) and (16) naturally describe different situations despite the fact that they are structurally similar, i.e. they differ only with respect to tense and the lexical information in the context of the reciprocal expression *each other*. Sentence (15) most naturally receives a strong interpretation in which *each other* means SR such that every man knows every other man, while the interpretation of sentence (16) (and hence the meaning of *each other*) is likely to be weaker, such that each man bites only one other man.

Dalrymple et al.'s SMH can correctly account for this intuitive distinction, by making plausible contextual assumptions. The SMH selects one meaning for each occurrence of the reciprocal (among those that they suggest in their inventory of logical operators), namely the one which results in the strongest sentence meaning that is still consistent with context. To illustrate: *each other* in sentence (15) most likely means SR, because there are no contextual restrictions on the number of possible acquaintances that any of the three persons may have, and SR results in the strongest sentence meaning possible. Thus, the analysis predicts that (15) can only be used truthfully in *S6*, which is the only situation that SR categorizes for (15). By contrast, the SMH predicts that the reciprocal in (16) has a weaker meaning, precisely because it is restricted by context. Plausibly, each individual has only one mouth and can hence bite only one person at a time. If we take this restriction on the number of patients an agent can bite simultaneously as part of the context that affects reciprocal meaning, then the SMH cannot select SR for the reciprocal expression in (16). Selecting SR would result in *S6*, where every man bites two other men simultaneously, which is inconsistent with this contextual restriction. According to Dalrymple et al., the meaning of the reciprocal is weakened as far as context requires it to. For (16), the SMH selects the operator that results in the strongest reciprocal meaning among those that do not contradict our knowledge about three people biting each other. That weaker operator predicts (16) to be true in *S3*.[7] The SMH thus structurally takes into account non-logical, contextual information in predicting logical interpretation, with lexical information being a key element of this context. This is in great contrast to standard formal semantics, which is often blind to 'non-logical', lexical meanings such as of content words *know* and *bite*.

Nature of the SMH

What is the nature of the SMH as a meaning principle? Dalrymple et al. convincingly show that the variation in reciprocal meaning cannot be due to general pragmatic reasoning. In other words, it cannot be due to the possibility of speaker meaning simply being stronger than literal meaning (for example due to conversational implicatures (Grice, 1975)) and it also cannot be due to the possibility of speaker meaning being weaker than literal meaning (due to loose speech). Instead, they show that what varies is the literal meaning of the reciprocal sentence. A crucial example that they use to support their claim is in (17).

(17) House of Commons etiquette requires **legislators** to address only the speaker of the House and **refer to each other indirectly**.

Sentence (17) requires that each legislator addresses every other legislator indirectly (Dalrymple et al., 1998, p.165). That means that it is true only in that case, and false otherwise. Dalrymple et al. show that this strong meaning (where the reciprocal means SR) cannot come from pragmatic strengthening of a fixed weak meaning, since exceptions are not allowed (example (18)), unless it is explicitly stated that they are exceptions (example (19)).

(18) Legislators refer to each other indirectly. #The most senior one addresses the most junior one directly.

(19) Legislators refer to each other indirectly, except the most senior one addresses the most junior one directly.

Thus, postulating fixed weak truth conditions for the reciprocal and assuming some general mechanism of pragmatic strengthening on top of it cannot be the right account.

On the other hand, postulating fixed strong truth conditions, which may be weakened by general pragmatic principles, also does not seem to work. Consider

[7] At this point, it does not matter exactly which weaker operator the SMH selects, as (16) is true in *S3* according to a couple of them.

example (20). Firstly, example (20) has a weaker meaning than example (17) above, since it does not require that each pitcher sat alongside every other pitcher.

(20) Five Boston pitchers sat alongside each other

The crucial point, however, is that this weakening of truth conditions is not arbitrary. Obviously, the meaning of (20) is weak in the sense that it does not require a pitcher to sit alongside more than two other pitchers, but moreover, in situations that are even weaker (e.g. where pitchers are separated by non-pitchers), the sentence is strictly false (p. 166).

Dalrymple et al. conclude that “the truth of the different reciprocal statements depends on meeting conditions of varying strength” (p.167). Neither postulating fixed weak truth conditions and assuming pragmatic strengthening nor postulating fixed strong truth conditions and relying on pragmatic weakening are adequate for the given data. Some of the reciprocal sentences that they present are literally false if the stronger conditions are not met (e.g. (17)), but on the other hand, some reciprocal sentences “can be strictly true even in conditions falling far short of those required by Strong Reciprocity”, e.g. (20). They conclude that “the literal meaning of the reciprocal really does vary across statements” and that we must have “a semantic explanation of how it varies; we cannot claim that the literal meaning is fixed and only the speaker’s meaning varies” (p. 167). Their explanation is the SMH.

Empirical value of the SMH and its limitations

The SMH, as Dalrymple et al. present it, has a large empirical benefit. In various cases it successfully predicts which of the various meanings a reciprocal expression will have, given the context in which it appears. Thus, the SMH is a satisfactory account, if we assume that the context indeed does what the SMH takes it to do.

However, as nice as the intuition itself is, the SMH is hard to test. The gist of the SMH is dependency on context when determining interpretation, although the authors do not delve too much into what exactly constitutes ‘context’. Without a specific notion of context, it is very difficult to empirically support the SMH, since it may involve many different variables (e.g. time of utterance, place of utterance, speaker intention). An important question motivating this dissertation, is what the relevant contextual information that needs to be taken into account actually is. For the contrast between (15) and (16), and in fact for all cases that Dalrymple et al. discuss, their analysis relies on the assumption that the context categorically restricts the number of patients that an agent may have simultaneously. This is a strong logical restriction on verb concepts such as *bite* vs. *know*: whether two simultaneous patients are physically possible for a given agent, i.e. categorized as an instance of the concept. Now consider again example (5), repeated as (21).

(21) John, Bill and George are pinching each other

According to the SMH, sentence (21) can only be used truthfully in a situation where every man pinches every other man (*S6*). The reasoning is as follows. In normal contexts, any agent is physically able to pinch two patients simultaneously. Hence, there are no strong logical restrictions on the verb concept *pinch* in (21). As a result, the SMH selects SR as the meaning of the reciprocal in (21). This is the strongest candidate meaning that does not contradict context. All weaker meanings are consequently disallowed. Thus, the only situation that SR categorizes for (21) is *S6*. This is problematic, however, as we will see later on in this chapter that in a preference task, (21) is often preferred in *S3* over *S6*, and in an independent truth-value judgment task, (21) is highly acceptable in *S3*. I will claim that any proposal accounting for reciprocal meanings needs to take into account more than sharp conceptual distinctions alone. In particular, I claim that more precise conceptual information in terms of typicality information is relevant. The challenge for the SMH and how typicality information can help to explain the facts will be discussed in more detail in section 2.3.

Sabato and Winter (2012): Maximal Interpretation Hypothesis
Following Dalrymple et al., Sabato and Winter (2012) also propose an explanation of the variation in reciprocal interpretations. In contrast to Dalrymple et al., they assume that not all contextual information is relevant for predicting reciprocal interpretation. In an earlier work, Sabato (2006) gives a counter example, repeated here as (22).

(22) #John doesn't know Bill and Bill doesn't know John. John, Bill and Jim know each other.

The first statement in (22) is part of the context of the second statement. According to the SMH, the interpretation of the second statement should therefore be affected by the first. Specifically, its interpretation should be weakened as far as the context of the first statement pushes it to. In other words, (22) should be fine, with the second statement having a weak interpretation in which John and Bill don't know each other. However, we generally consider (22) to express a contradiction.

Since clearly not all context appears to be relevant, Sabato and Winter (2012) zoom in on one aspect of that context that they believe captures the facts: the semantic restrictions on the denotation of *R*, the predicate in the scope of the reciprocal. Whereas the SMH consisted of two components, namely an inventory of possible meanings and a selection principle that picks out one of those meanings, Sabato and Winter's Maximal Interpretation Hypothesis (MIH) consists of only the selection component. The authors propose that the interpretation of each reciprocal can be predicted immediately from the properties of the predicate *R* in its scope, since each predicate has its own inherent properties which a priori limit the range of possible reciprocal interpretations. Their proposal works as follows. Each relational predicate *R* has a domain of interpretations Θ_{REL}, containing all relations (or 'graphs') that are possible as denotations of the predicate, based on its semantic restrictions. The MIH then selects the reciprocal interpretation from Θ_{REL} that is not properly contained in any other relation in Θ_{REL} – i.e. the maximal relation within Θ_{REL}. To illustrate the proposal, consider (23) and (24).

(23) The boys like each other
(24) The students gave each other measles

According to the MIH, the different interpretations of (23) and (24) arise purely due to the different semantic restrictions that are inherent to the predicates *like* and *give measles to*. The predicate *like* has no semantic restrictions, meaning that any binary relation is in its reciprocal interpretation domain Θ_{LIKE}. The maximal relation within Θ_{LIKE}, one that is not properly contained in any other relation in Θ_{LIKE}, must therefore be SR. By contrast, the predicate *give measles to* is restricted such that it can only be the inverse of a function (a person can only get measles once) and it can only denote acyclic relations (measles cannot be passed around in a circle) (Sabato, 2006, p.33). As a result, the domain $\Theta_{\text{GIVEMEASLESTO}}$ is smaller since it only contains those relations that are consistent with these restrictions. The MIH then selects the maximal one among them, which is one in which "every student is connected to every other student by the transitive and symmetric closure of the denotation of *give measles*" (Sabato, 2006, p. 34).

To summarize, the MIH is similar to the SMH in selecting one candidate meaning for each occurrence of the reciprocal based on maximality. The difference is in the way maximality is determined. The SMH selects the strongest possible meaning among a set of meanings that were attested in natural language, while the MIH determines the maximal interpretation directly from the inherent properties of the predicate *R*, without using an independent set of meanings.

Limitations of the MIH

Despite the fact that the SMH and the MIH are substantially different, the MIH is sensitive to similar problems as the SMH. It defines 'context' in a much more controlled manner, but still fails to account for examples that were problematic for the SMH such as (21). The predicate *pinch* in (21) has a maximal interpretation domain identical to the predicate *like* that was discussed above for (23), since there are no semantic restrictions (at least not in the case of three individuals). The predicted interpretation of the reciprocal expression in sentence (21) is thus still SR, since this is the maximal relation within Θ_{PINCH}. Thus, again, this hypothesis is unable to make the right prediction for (21).

Mari (2013)

Mari (2013) specifically focuses on sentences that satisfy the truth conditions of the operator IAO (see subsection 2.2.1) and proposes an explanation as to why some of these sentences are felicitous while others are not. Consider sentences (25) and (26).

(25) Susan and Mary followed each other into the elevator
(26) #Susan and Mary gave birth to each other

Both (25) and (26) could in principle satisfy the truth conditions of the operator that Dalrymple et al. (1998) called IAO. Dalrymple et al. claim that IAO can be obtained if a sentence contains an asymmetric relation. However, while both (25) and (26) contain asymmetric relations (*follow*, *give birth to*), only (25) is felicitous. With challenging examples such as (26), Mari (2013) points out that the truth conditions for IAO are not generally available for reciprocal sentences with asymmetric relations, hence proposals like the SMH overgenerate. The author claims that determining "where the break point stands between sentences that satisfy IAO but are false (e.g. 26) and those that satisfy IAO and are true (e.g. 25) will put us on the track of a new temporal-modal account". She proposes an account in which SR has a privileged role, such that reciprocal sentences are "true and acceptable if and only if the relation is either actually or possibly strong reciprocal on the reference set, provided that the possibilities are reasonable". To define what is 'reasonable', she uses a branching-time framework. Within that framework, (26) is predicted to be infelicitous because SR cannot hold in any reasonable future in which the sentence is evaluated – unlike (25). For an elaborate explanation of this account, see Mari (2013). This account, while providing important insight into temporal considerations that are relevant for reciprocal sentences, is not immediately relevant for the work presented in this chapter. In fact, a sentence like (21) is still problematic since the proposal that Mari puts forward would predict it to mean SR, as there is no logical asymmetry that the predicate *pinch* forces: if A pinches B, B may or may not be pinching A at the same time.

Interim summary

This section has reviewed proposals that aim to explain why different reciprocal sentences receive the interpretations that they do. Dalrymple et al.'s SMH selects the strongest candidate meaning for each reciprocal expression from an inventory of attested meanings, based on the context in which it occurs. Counter examples such as (21) for which the SMH makes the wrong predictions point out that there is a problem with this hypothesis. I believe that the main issue is not the gist of their proposal, but the fact that context is not defined, likely due to the lack of experimental data in their work (as well as in the works of others drawing on it such as Beck, 2001; Sabato and Winter, 2012; Mari, 2013).

Sabato and Winter (2012) make an attempt at restricting the context that the reciprocal is sensitive to, to the semantic restrictions on the predicate *R* in the scope of the reciprocal. Despite this, their MIH also fails to capture the interpretation of the problematic example in (21).

Finally, I briefly discussed Mari's (2013) account which focuses on reciprocal sentences that satisfy IAO. Based on the behavior of this type of sentences, Mari formulates a temporal-modal account of reciprocal sentences in which SR has a privileged role**.** Similar to the SMH and MIH, Mari's proposal has trouble accounting for (21).

2.3 The Maximal Typicality Hypothesis

This section introduces the Maximal Typicality Hypothesis as a solution to the problems that were described for previous proposals, given the acceptability patterns for sentences (15), (16) and (21) – repeated below as (27)-(29).

(27) John, Bill and George know each other
(28) John, Bill and George are pinching each other
(29) John, Bill and George are biting each other

We[8] argue that these problems appear because there are certain aspects of meaning that those proposals do not take into account. We propose a method for specifying a central contextual parameter that affects the interpretation of reciprocal sentences. This method is incorporated in a new principle, the Maximal Typicality Hypothesis (MTH), which is based on initial work by Kerem, Friedmann and Winter (2009). After some background on typicality in subsection 2.3.1, the MTH is introduced in section 2.3.2. Further, subsections 2.3.3-2.3.6 explain the method for testing the MTH and describe previous work by Kerem, Friedmann and Winter (2009).

2.3.1 Concept typicality

In terms of theories of mental concepts (Margolis & Laurence, 1999), previous proposals only take into account 'sharp' aspects of the meaning of verb concepts such as *know*, *bite* and *pinch*: whether certain situations – e.g. those compatible with SR – are possible or impossible in a given context. Such sharp distinctions do not account for the experimental contrasts we find between sentences like (27), (28) and (29). To overcome this shortcoming, Kerem, Friedmann and Winter (2009) proposed that

[8] The work described in sections 2.3 and 2.4 was carried out together with Marijn Struiksma and Yoad Winter.

typicality effects are a critical part of the contextual information that affects reciprocal interpretations.

Since the 1970's, a host of psychological studies has shown that subjects consistently rank some instances of a one-place concept as more typical than others (e.g. Rosch, 1973; Smith, Shoben & Rips, 1974; Rosch & Mervis, 1975). For example, besides being able to categorize sparrows and ostriches within the *bird* category, and bats and koalas outside of it, subjects also distinguish between members of a category: e.g. when people are asked to rank bird instances, sparrows are judged as more typical for the concept *bird* than ostriches. Such rankings correlate with other measures of typicality, such as categorization speed (more typical instances are categorized faster than less typical ones) and error rate (more typical instances lead to fewer categorization errors than less typical ones). Throughout this dissertation, the term 'typicality effects' is used to refer to these basic behavioral phenomena about categorization.

Moreover, I follow Hampton (2007) in assuming that deciding on an instance's category membership and judging its typicality as an instance of that category are two related behavioral measurements, based on one and the same underlying variable. For example, there is a correlation between binary membership measures for *sparrow* (1), *ostrich* (1), *bat* (0) or *koala* (0) on the one hand, and their typicality rating on the other hand (*sparrow* > *ostrich* > *bat* > *koala*). Hampton assumes a so-called threshold model, according to which there is a threshold somewhere along a typicality function that makes a binary distinction between members (*sparrow*, *ostrich*) and non-members (*bat*, *koala*).

While many nouns categorize entities, verbs categorize situations: events and states containing one or more entities as participants. As shown in our experimental work, verb concepts such as *pinch* exhibit typicality effects with situations, similarly to the typicality effects that noun concepts such as *bird* show with entities. In other words, people can consistently rank some instances of the concept *pinch* as more typical than others. For reciprocal sentences and verb concepts, taking typicality into account means changing perspectives about the previously used notion of context. In addition to the definitional aspects that proposals like the SMH consider (can a given situation be categorized as an instance of a verb concept *X*?), we also have recourse to aspects of typicality (what preferences between situations does a verb concept *X* induce?). This allows us to take into account more factors that affect the interpretation of reciprocal sentences than in Dalrymple et al.'s or Sabato and Winter's proposals. As we saw, the only distinction that the SMH and MIH make – i.e. between possible and impossible situations - forces them to choose SR as the meaning of the reciprocal expression in both sentences (27) and (28). This is incorrect for (28). Under a theory that takes into account typicality effects - making a further distinction within possible situations - the interpretation of the two sentences differs due to differences in typicality information between the concepts *know* and *pinch*. Kerem, Friedmann and Winter (2009) first proposed such a theory, which uses experimental evidence on typicality to make predictions about the interpretations of reciprocal sentences. According to that theory, "a reciprocal expression requires the denotation of its predicate antecedent to be a relation of maximal typicality relative to the predicate concept" (Kerem, Friedmann & Winter (2009), p. 4).

In the next sections, I introduce our main hypothesis: the Maximal Typicality Hypothesis (MTH). The MTH is based on Kerem, Friedmann and Winter's hypothesis, but unlike their hypothesis contains the process of selecting a so-called 'core situation' which we use to make precise predictions about the interpretation of reciprocal sentences in various situations. Next, I will discuss Kerem, Friedmann and Winter's initial ground work, and finally continue with our comparative experimental study of the predictions of our MTH as opposed to the SMH and MIH.

2.3.2 The MTH: Connecting typicality with the interpretation of reciprocals

The MTH is a meaning principle that uses information about typicality of different situations with respect to a given verb concept *P*. On the basis of this information on *P*, the MTH singles out one situation as the *core situation* that is described by a reciprocal sentence containing *P*. Intuitively, we describe the core situation as the situation that contains "enough relations" between the agents to fully satisfy reciprocity.

In terms of reciprocity, for sentence (27) with the verb *know*, only situation *S6* has enough relations, whereas for sentences (28) or (29) with the verbs *pinch* or *bite*, *S3* already has enough relations. To be more precise, we need to define what we mean by "enough for reciprocity", and how the sentences differ in this respect. According to the MTH, the differences between the sentences follows from the observed typicality difference between the verb concepts. (These differences in typicality are measured experimentally, as we will see in section 2.4.) With the concepts *pinch* and *bite* we cannot add relations to *S3* without reaching a situation where one agent pinches/bites two patients. Such a situation would be atypical for the verb concepts. And because we cannot add relations to *S3* without a reduction in its typicality for the concept *pinch* or *bite*, we consider *S3* to have "enough" relations for sentences (28) and (29). By contrast, with the verb *know*, we can add relations to *S3* without any change in typicality for the verb concept. Thus, *S3* does not attain "enough" relations for (27). In technical terms, we describe this difference by observing that *S3* is the *maximal situation* among the most typical situations for the verb concepts *pinch* and *bite*. By contrast, *S3* is not maximal among the most typical situations for the verb *know*. Accordingly, the MTH selects *S3* as the core situation for sentences (28) and (29), but not for sentence (27). In general, this process of selecting the core situation on the basis of typicality information that is conveyed by the main concept is defined below.

Maximal Typicality Hypothesis (MTH): For a reciprocal sentence with a verb concept P in the scope of the reciprocal expression, situation S_C is the *core situation* for the sentence iff S_C is maximal among the situations that are most typical for P. [9]

The MTH assumes that similar to noun concepts, verb concepts invoke typicality judgments.[10] Once we know which situations are most typical for a given verb concept, the MTH predicts the core situation for a reciprocal sentence containing that verb. The key for selecting the core situation among the most typical situations is the notion of 'maximal situation'. This notion relies on our ability to order situations according to containment relations between them. For example, a situation with three individuals in which each individual pinches two individuals (*S6*) properly contains any situation in which each of those individuals pinches only one of those individuals (*S3*). To illustrate this notion of containment further, let us consider the diagrams in Figure 1.

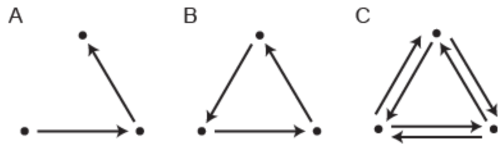

Figure 1: Three situations with three agents

Figure 1 schematically represents three different situations. Each of these situations includes three agents, represented by dots. The arrows represent directed actions that are categorized by a particular relational concept like *pinch*. Figure 1A represents a situation in which two of the three individuals act on one other individual. We will refer to this kind of situations as *S2*. Figure 1B represents a situation in which each of the three individuals acts on one other individual and is acted on by one other individual.

[9]Formally, the MTH defines the core situation as any maximal situation S_C among the situations S that attain a local maximum of the function $TYP_P(S)$ – the typicality of S for the predicate P. In principle, there may be two or more situations that attain such a maximum, but without any of these situations containing the other. In such cases, the MTH defines more than one situation as a core situation. However, in ordinary language use it is not easy to come up with such cases, and they do not surface in our experiments. For this reason, here and henceforth we assume uniqueness of the core situation, and refer to it as '<u>the</u> core situation'.

[10] Verb concepts like *pinch* categorize events while verb concepts like *know* categorize states. The event/state distinction is likely to affect typicality preferences with verbs, hence the predictions of the MTH. However, since it does not directly affect the formulation of the MTH, we will not focus on it here. A similar point holds for tense: event verbs like *pinch* can take the progressive tense, while state verbs like *know* cannot. The choice of tense naturally affects typicality preferences. However, as we shall see, the MTH is about the correlation between typicality preferences and reciprocal interpretation. For this reason, holding the tense variable constant in the two measures does not affect the predicted correlation between them.

We have referred to such situations as *S3*. Figure 1C represents a situation in which each of the individuals acts on both of the other individuals. Such situations have been referred to as *S6*. Situation *S6* properly contains *S3*, which properly contains *S2*. Accordingly, among these three situations, *S6* is maximal, since it contains both *S3* and *S2*, and is not contained by any of them. In sentence (27), where we assume that *S6*, *S3* and *S2* are of equal typicality for the verb *know*, *S6* is the maximal one among them, hence it is selected by the MTH. By contrast, in sentences (28) and (29), we assume that *S2* and *S3* are most typical for the verbs *bite* and *pinch*, whereas *S6* is less typical (or even impossible, with *bite*). The maximal situation among the most typical situations is now *S3,* and accordingly, this is the situation that is selected by the MTH as the core situation.

Now that we have defined the most intuitive, 'core' situation for a reciprocal sentence using the notion of containment, we use the MTH for explaining the acceptability pattern that we observe for reciprocal sentences such as (27)-(29) in all situations, including those that are not the core situation. Firstly, reciprocal sentences are always expected to be highly acceptable in the core situation that the MTH selects for it. Secondly, there are two possible types of situations that differ from the core situation:

a) *Situations that are properly contained in the core situation* (i.e. those that contain fewer relations): in those situations, reciprocal sentences are expected to be less acceptable than in the core situation, with decreasing acceptability the fewer relations there are. For example, if the MTH selects *S6* as the core situation for a reciprocal sentence (e.g. sentence (27)), then that sentence is predicted to be less acceptable in *S3* than in *S6*, and less acceptable in *S2* than in *S3*.

b) *Situations that properly contain the core situation* (i.e. those that contain more relations): in those situations, there are "more than enough" relations to support reciprocity. Whether such situations are relevant for actual use of reciprocal sentences is determined by whether speakers categorize such situations as possible instances of the relevant verb concept. For example, if *S3* is selected as the core situation for both sentences (28) and (29), then both sentences are predicted to be highly acceptable in this situation. However, as for *S6* situations, judgments are affected by whether *S6* is judged as a possible instance of the verb concepts *pinch* and *bite*. Situation *S6* is a possible instance of the concept *pinch*. Accordingly, speakers are expected to have judgments on sentence (28) in *S6*, and to judge the sentence as highly acceptable. By contrast, *S6* situations are unlikely to be judged as instances of the concept *bite* to begin with. Accordingly, sentence (29) will be much less acceptable in *S6*, not because it is false, but because speakers are expected not to have judgments on (29) in *S6*.

2.3.3 Approximating typicality of complex situations

In order to test the MTH, we first need information regarding the typicality of situations like *S2*, *S3* and *S6* for different verb concepts. The reason we consider situations like *S6* as atypical for the verb *pinch* is because agents in it pinch more than one patient simultaneously. Using this kind of typicality information, the predictions of the MTH are tested against acceptability judgments on reciprocal sentences in situations like *S2*, *S3* and *S6*. But how can we systematically gather experimental information about the typicality of situations like *S2*, *S3* and *S6* for different verb concepts?

As was discussed in subsection 2.3.1, previous works have measured typicality effects using standard tasks such as categorizing instances or ranking them with respect to how typical they are. Those works have mostly measured typicality effects for concepts expressed by nouns. For situations like *S2*, *S3* and *S6*, directly measuring typicality effects in such a way is a more difficult task. Situations like those that we would like to test involve a couple of agents, a couple of patients, and a couple of relations between them. Thus, comparing *S3* to *S6*, for example, means comparing more than just one aspect. Such a multitude of differences between the compared situations make the task more difficult, potentially creating more divergence among speakers. Consider for example the two situations in Figure 2, which depict two possible situations with the verb concept *pinch*.

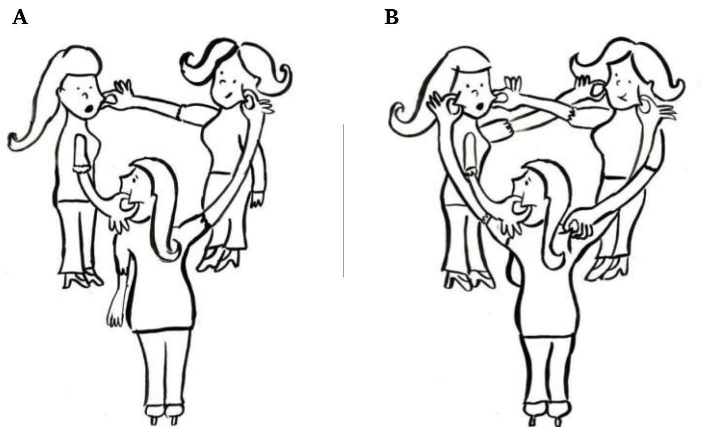

Figure 2: Two possible situations for the verb concept *pinch*: *S3* (A) and *S6* (B).

Both situations show three women in pinching activities. In situation A, every woman is pinching only one other woman (*S3*), while in situation B, every woman is pinching every other woman (*S6*). In its account of reciprocal sentences, the MTH relies on the assumption that situations as illustrated in Figure 2 may differ with respect to how typical they are for the verb concept *pinch*. Thus, the experimental difficulty to test typicality effects directly with such situations is a challenge for testing the MTH and its potential improvement over the SMH and other proposals.

To tackle this challenge for situations like *S2*, *S3* and *S6*, we approximate typicality by looking at their sub-situations in relation to requirements of the verb. These sub-situations differ only in one, basic respect, creating an easier task. In Figure 2, situation B is intuitively expected to be less typical for *pinch* than situation A because the agents are pinching with both hands. We test this expectation with a simplified measurement for the typicality of different situations as in Figure 2. That measurement tests a basic parameter that distinguishes those situations: how many patients are simultaneously acted upon by a given agent. We refer to this parameter as *patient cardinality*. For instance, in Figure 2, we describe the difference between the two situations as a simple difference in patient cardinality: one per agent (Figure 2A) vs. two per agent (Figure 2B). This reduces the comparison between the two situations in Figure 2 to a comparison between the two situations in Figure 3. When using simpler situations as in Figure 3, we can more easily measure which is a more typical instance of the verb *pinch*. If Figure 3A is preferred over Figure 3B, we infer that *S3* situations (Figure 1B) are more typical for the verb concept *pinch* than *S6* situations (Figure 1C). This information is used for evaluating the MTH.

A **B**

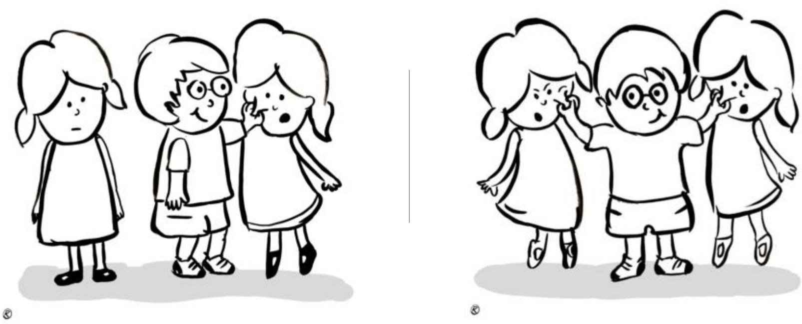

Figure 3: Two instances of the verb *pinch*: a one-patient situation (A) and a two-patient situation (B).

2.3.4 Predictions of the MTH

Based on the measurement of typicality effects described above in 2.3.3, let us now illustrate the precise empirical predictions of the MTH for the interpretation of sentences (27)-(29). The typicality test of Experiment 1 (described in detail in section 2.4) shows that for the verb *know* in (27), there is no typicality preference between simple situations with different patient cardinality. This means that there is no difference in the representativeness of different instances of *knowing* for the concept, in terms of how many patients each agent knows. For instance, a state in which a person knows one other person is just as typical an instance of *know* as a state in which a person knows two other people. Using this measurement, we extrapolate that there is also no difference in typicality between reciprocal situations that differ merely in terms of patient cardinality (at least in terms of one patient vs. two). This means that the reciprocal situations *S2*, *S3* and *S6* are equally typical for the verb concept *know*. With this extrapolation about typicality of candidate situations, the MTH predicts that the core situation that (27) describes is *S6*: the maximal situation among the three situations *S2, S3* and *S6*. We thus predict that (27) is acceptable in *S6*. This agrees with the predictions of the SMH and the MIH that the reciprocal expression in (27) means SR. Furthermore, we predict (27) to be less acceptable in *S3*, and even less so in *S2*, because these situations are contained in the core situation *S6*. These are more fine-grained predictions than the mere truth-falsity predictions of the SMH and MIH.

In contrast to the verb *know*, the verb *bite* in (29) does show a typicality effect with respect to patient cardinality. In particular, instances of the concept *bite* that have one patient per agent are judged as being more typical than instances that have two patients per agent simultaneously.[11] We extrapolate from this that in the most typical reciprocal situations for the verb *bite*, there is no more than one patient per agent. The MTH predicts that the core situation described by sentence (29) is the maximal one among those situations. This is the situation in which each agent bites exactly one patient: *S3*. Adding more *biting* relations to such a situation would result in a situation that is not among the most typical situations. Thus, we predict (29) to be acceptable in *S3*, in line with the predictions of the SMH and MIH. Furthermore, we predict (29) to be unacceptable in *S6*, since despite the fact that it is basically true in *S6*, *S6* is unlikely to be judged a possible instance of the concept *bite*. Finally, sentence (29) is expected to be less acceptable in *S2* than in *S3*, because *S2* is properly contained in the core situation.

The example that most clearly distinguishes the predictions of the MTH and previous proposals is sentence (28). In this case, the core situation that is specified by the MTH does not coincide with the strongest possible interpretation that is selected by the SMH or MIH. As we saw, the SR meaning, where every boy is pinching every other boy, is physically possible for sentence (28). Therefore, according to the SMH and MIH, SR is the only possible reading for the reciprocal expression, and *S6* is expected to be the only situation in which sentence (28) is true. By contrast, the MTH predicts (28) to be true in *S3*. As we will experimentally show, the concept *pinch* shows the expected typicality effect with respect to patient cardinality: a situation in which one agent pinches one patient is a more typical instance of *pinch* than a situation in which one agent pinches two patients simultaneously. In this sense, the verb *pinch* is similar to the verb *bite*, and we again extrapolate that the most typical situations for the verb *pinch* are those in which there is at most one patient per agent. From these situations,

[11] In fact, the latter situation may even be considered physically impossible. This means that situations with multiple patients per agent may not be instances of the verb concept *bite*.

the MTH selects the maximal one, *S3*, as the core situation that the sentence describes. Accordingly, we expect sentence (28) to be fully acceptable in *S3*. In addition, since situation *S6* properly contains the core situation *S3*, we expect the sentence to be true in *S6*. Unlike *bite* in (29), the concept *pinch* allows *S6* as a possible instance of the concept, and hence we predict (28) to be acceptable in *S6*. The expectation that (28) is acceptable in both *S3* and *S6* distinguishes the predictions of the MTH from those of the SMH and MIH.

In order to test these predictions, we tested whether and how typicality preferences in terms of patient cardinality predict reciprocal interpretations using the MTH. As we will show, judgments on sentences like (28) reveal a systematic advantage of the MTH over the SMH and other works.

2.3.5 Kerem, Friedmann & Winter (2009)

Kerem, Friedmann and Winter (2009) first investigated the relationship between typicality effects with verb concepts and the interpretation of reciprocal sentences with Hebrew speakers. Any study aiming to test that relation requires two kinds of information, hence pairs of experiments - one that tests reciprocal interpretation and one that independently tests typicality effects – plus a correlation analysis that tests their relation. Kerem, Friedmann and Winter conducted two of such experiment pairs, a pictorial pair and a textual pair. While the pictorial pair restricted the verbs that could be tested to depictable action verbs such as *pinch* and *bite*, the textual pair made it possible to also test stative verbs that cannot easily be depicted, such as *know* and *hate*.

Kerem, Friedmann and Winter's pictorial typicality test was a forced-choice task that measured 32 different Hebrew verbs with respect to their preference for a particular instance of that verb concept, specifically a one-patient situation vs. a two-patient situation. This way they aimed to measure which of these two situations is a more typical instance of the verb concept, in order to extrapolate which reciprocal situations (*S2*, *S3* and *S6*) are more typical than others – as was explained before. An example of a test item (for the verb *stab*) is given in Figure 4.

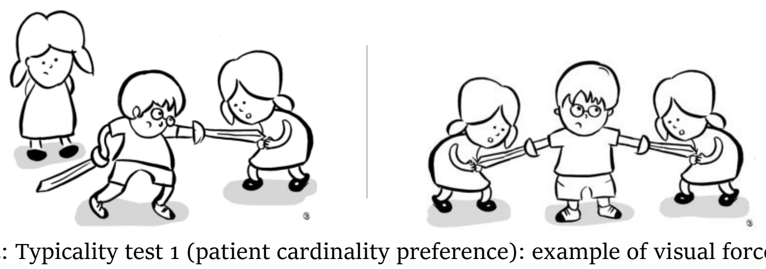

Figure 4: Typicality test 1 (patient cardinality preference): example of visual forced-choice task for *to stab* (Kerem, Friedmann & Winter, 2009).

All test items contained a verbal description of the situation in Hebrew: an agent-only sentence referring to an activity associated with the verb concept, for example the Hebrew correlate of “The boy is stabbing”. Participants were instructed to indicate which of the two depicted events better describes the sentence.[12]

The textual typicality test was also a forced-choice task, measuring 11 different verbs on their preference with respect to a particular instance of that verb concept. Participants were given incomplete transitive sentences in which the object noun phrase was missing. They were instructed to complete the sentence in a way that sounded best to them, with a choice between a singular object and a plural object – reflecting the same choice as the pictorial test, i.e. one-patient situation vs. two-patient situation. Two examples, translated from Hebrew, are in (30).

(30) Typicality test 2 (patient cardinality preference): examples of textual forced-choice task for *to stab* (Kerem, Friedmann & Winter, 2009).

(*perfective*)	Last year, Athos stabbed…	a. the girl	b. the girls
(*imperfective*)	Rina entered and noticed Athos stabbing…	a. the girl	b. the girls

Four of the verbs were only tested in the perfective aspect (since the imperfective is not applicable for those verbs, e.g. *hate*, which cannot be used in ?*Rina entered and noticed Athos hating the girl(s)*), while seven of them were tested in both the perfective and imperfective aspect (as in (30)).

For the pictorial interpretation test, Kerem, Friedmann and Winter selected 13 verbs from the pictorial typicality test and presented them in a Hebrew reciprocal sentence, e.g. the correlate of “Athos, Portos and Aramis are stabbing each other”. In these main items, participants were presented with two realistically depicted situations with human agents: a situation in which every individual acted on every other individual (*S6*) vs. a situation in which every individual acted on only one other individual (*S3*). The example situations for “Athos, Portos and Aramis are stabbing each other” are given in Figure 5. Subjects were again given a forced-choice task, and were now instructed to indicate which of the two depicted situations better describes the sentence.

[12] The authors claim that this phrasing proved clearest for their participants, even though it might not be perfect from a theoretical semantic point of view.

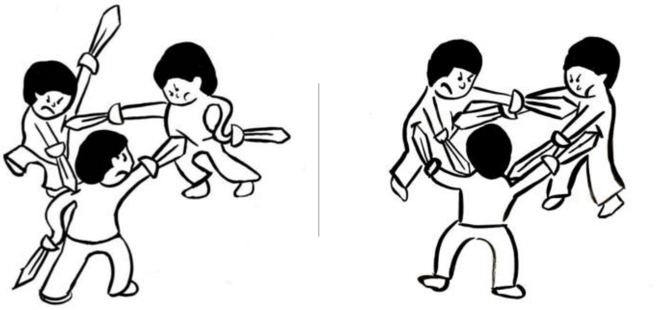

Figure 5: Interpretation test 1: example of visual forced-choice task for *Athos, Portos and Aramis are stabbing each other* (Kerem, Friedmann & Winter, 2009).

The textual interpretation test was an inference task that measured the interpretation of reciprocal sentences containing the same predicates from the textual typicality test. Participants were presented with reciprocal sentences with subjects that referred to three individuals. They were then instructed to indicate whether it is necessary or not to conclude from the sentence that one, randomly chosen individual from the three (e.g. Athos) acted on another randomly chosen one from the three (e.g. Aramis). Examples are in (31), translated from Hebrew. The reasoning was that a positive answer indicates Strong Reciprocity, with all possible six pairs among the three individuals, while a negative answer indicates a weaker meaning.

(31) Interpretation test 2: example of textual inference task for *Athos, Portos and Aramis are stabbing each other* (Kerem, Friedmann & Winter, 2009).

(*perfective*)
Last year, Athos, Portos and Aramis stabbed each other
Can you deduce from this that Aramis stabbed Portos?

(*imperfective*)
Rina entered and noticed Athos, Portos and Aramis stabbing each other
Can you deduce from this that Aramis stabbed Portos?

Results of the typicality tests showed different preferences regarding patient cardinality for different verbs. In the pictorial test, the proportion of participants who preferred the drawing depicting a one-patient situation over the drawing depicting a two-patient situation ranged from 8% to 90%. In general however, most of the verbs showed a preference for the one-patient situation, indicating that they considered the situation in which only one patient was involved as a more typical instance of that verb concept than the situation with two patients. In the textual test, the preference for one-patient situations (singular object) ranged from 10% to 96%.

Results of the interpretation tests also demonstrated differences between items. Specifically, the preference pattern for the two situations differed across verbs. In the pictorial interpretation test, the proportion of participants who preferred the *S3* situation over the *S6* situation ranged from 19% to 67%. Crucially, for eight out of the thirteen reciprocal sentences that were tested, more than one third of the participants preferred *S3* over *S6*. This is unexpected under the SMH, which predicts that the meaning of any reciprocal sentence is the strongest one that is consistent with its context. In the experiment, the relevant restrictions that the context induces were controlled by making sure that situation *S6* is not considered impossible when judging the sentences, i.e. it was realistically depicted. In this kind of context, the SMH would predict reciprocal sentences to be true in the stronger situation *S6* and false in *S3*. Thus, the SMH predicts 100 percent preference for *S6* over *S3*. This expectation was not borne out: many subjects preferred *S3* over *S6* as the best situation for the reciprocal sentence. This preference pattern is also unexpected under the MIH, since it would select as the one meaning of the sentence that meaning that is maximal given the verb's interpretation domain – which was SR for all tested verbs, which only holds in *S6*. In the textual interpretation test, the proportion of participants who preferred *S3* over *S6* ranged from 0% to 41%.

The main result that Kerem, Friedmann and Winter (2009) were interested in, was the relation between the two measured variables: does typicality affect interpretation? Can typicality effects explain why participants at times prefer *S3* over *S6*, even though *S6* is realistically depicted and hence possible in the context of utterance? In other words, can we predict the results of part two from the results of part one? In order to determine this, they calculated the correlation between the measures, specifically the correlation between the preference for the one-patient situation in the typicality test and the preference for *S3* in the interpretation test. For the pictorial tests, they found merely a non-significant correlation of this type (r (12) = .42, p (one-tailed) = .089), after removing one outliner. In the textual test, they found a significant correlation (r (15) = .56, p (one-tailed) = .015), but only when analyzing merely the verbs that were tested in both aspects, and only after removing one of those seven verbs.

In sum, all we can definitively conclude from Kerem, Friedmann and Winter's work is that both the SMH and the MIH fail to make the correct predictions. This strengthens our assumption that these hypotheses should be replaced by another principle, which employs a more fine-grained analysis of binary concepts.

2.3.6 The current study

Our formulation of the MTH submits that the core situation for any reciprocal sentence is maximal among the situations that are most typical for the verb concept in the reciprocal's scope. While we can clearly conclude from Kerem, Friedmann and Winter's results that the SMH and MIH fail, we cannot yet strongly support the MTH based on these results alone. The way to support the MTH would be to show that typicality systematically correlates with reciprocal interpretation. However, the results from Kerem, Friedmann and Winter's pictorial experiment (2009) do not yield a correlation between the preference for *S3* and the preference for a one-patient situation. Plausibly, this is due to the fact that the distribution of values for patient cardinality preferences in their typicality tests did not vary to a high degree. With their pictorial experiment, one reason for that may be that the method only allowed testing action verbs, which in general turned out to have a preference for one-patient situations. Furthermore, their textual experiment had restrictions due to tense, i.e. some verbs were tested in the perfective aspect only while others were tested both in the perfective and the imperfective aspect.

Another, even more important reason why we cannot yet support our formulation of the MTH based on Kerem, Friedmann and Winter's results is due to their use of a forced-choice task. While this type of task was necessary for creating a context of utterance that allowed both *S3* and *S6* in order to argue against the SMH, testing the predictions of the MTH requires a different type of task. Our MTH makes predictions about acceptability of situations like *S3* and *S6*, not preferences between them. Thus, while the considerable preference rates for *S3* in Kerem, Friedmann and Winter (2009) are clear evidence against the SMH, we need truth-value judgments in order to provide conclusive evidence in favor of the MTH. As we saw, for some sentences, the MTH predicts high acceptance rates for both *S3* and *S6*, which cannot be measured though a preference task.

In sum, testing our MTH requires two additional things. Firstly, it requires various types of verb concepts in terms of their patient-cardinality preferences. Accordingly, our experiment included a wide selection of verbs. In order to test these all within the same paradigm, we used different visual presentation of the stimuli. Secondly, providing evidence for the MTH requires information regarding the acceptability of reciprocal sentences given particular situations. Therefore, the current study contained a truth-value judgment task on single situations instead of a forced-choice preference task between two situations.

2.4 Experiment 1: Testing the MTH with reciprocal sentences

2.4.1 Introduction

Experiment 1 studied Dutch[13] and tested the predictions of the Maximal Typicality Hypothesis by using schematic presentations of situations. This allowed us to compare different kinds of verbs in terms of their patient cardinality preferences (such as *know*, *bite*, and *pinch*) within one experiment, using exactly the same measures for each verb. This way, we avoid some of the limitations of both the pictorial and the textual presentation used in Kerem, Friedmann and Winter (2009).

In order to test the relation between typicality and reciprocal interpretation, we made use of two parts. The MTH is evaluated on the basis of a correlation analysis between the two parts. Part 1 of the experiment tested typicality effects with different types of verb concepts in isolation, specifically the preference between one-patient situations vs. two-patient situations. Part 2 of the experiment tested reciprocal interpretations, specifically whether reciprocal sentences are accepted in *S6* and *S3*. We included these situations because we expected both of them to be possible core situations for different sentences. *S6* was expected to be the core situation for sentences with verb concepts that show no patient cardinality preference. *S3* was expected to be the core situation for sentences with verb concepts that show a one-patient preference. In addition, we performed a correlation analysis to test the more fine-grained relationships that the MTH predicts, between the acceptability of *S3* in Part 2 and the preference for one-patient situations in Part 1. In this way, Experiment 1 tested the MTH account of the interpretation pattern in Kerem, Friedmann and Winter's experiment that the SMH and MIH do not capture.

Experiment 1 used schematic representations of situations, as in Figure 1 and Figure 6 (below). The verbs that were used are categorized into three types, based on their expected typicality effects regarding patient cardinality.[14]

---

[13] The reason that Dutch was tested was purely logistic: the location of the main research effort was the Netherlands. However, as mentioned below in subsection 2.4.2, Dutch and Hebrew show significant correlations regarding patient cardinality with verb concepts.

[14] Besides one-patient preference verbs and neutral verbs, Kerem, Friedmann and Winter (2009) also studied four verbs that showed a two-patient preference. However, these verbs were all collective verbs, e.g. *give a speech*, *photograph*. With such verbs, it is difficult to determine whether we can really speak of a preference for two patients or simply a preference for a larger collective patient over a smaller one – which makes sense for a verb like *give a speech*. With that verb, for example, Kerem, Friedmann and Winter's pictorial forced-choice test compared an image in which one speaker gives a single speech to one listener to an image in which one speaker gives a single speech to two listeners. This is substantially different from a verb like *pinch*, where the choice was between a person performing one action vs. two actions. In Experiment 1, we only added some collective verbs as fillers to the pretest, see subsection 2.4.2.

Type 1 – Neutral verbs: Verbs for which we expected no preference between instances with different patient cardinality (e.g. *kennen*, 'know').

Type 2 – One-patient-preference verbs: Verbs for which we expected a preference for situations with one patient, even though we expected two-patient situations to also be categorized as instances of the verb concept (e.g. *knijpen*, 'pinch').

Type 3 – Strong one-patient-preference verbs: Verbs for which we expected a high preference for one-patient situations, because we expected two-patient situations to not be categorized as instances of the verb concept at all (e.g. *bijten*, 'bite').[15]

This intuitive classification was used in the experiment as a means for selecting candidate verbs, as well as for testing large-scale typicality differences between groups of verbs. I will refer to these three types when discussing set-up and subresults of Part 1 and Part 2 of Experiment 2. Note however that the distinction between the three types is not meaningful in the final correlation analysis of all data points.

2.4.2 Pretest

In preparation for Experiment 1, we conducted a pretest in Dutch that measured only patient cardinality preferences. This pretest had several aims. Firstly, it tested whether measuring typicality effects with schematic stimuli is comparable to measuring them with pictorial stimuli as in Kerem, Friedmann and Winter (2009). Secondly, the pretest tested whether Dutch verbs behave in a comparable way to their Hebrew counterparts in Kerem, Friedmann and Winter's experiment. Thirdly, and most importantly, the results of the pretest were used for selecting the verbs to be tested in the main parts of Experiment 1. In Part 1 of the experiment, we aimed to obtain a wide range of cardinality preference values. This would ensure that the eventual correlation analysis with reciprocal interpretations is based on verbs that show diverse typicality effects. To this end, the pretest was used for selecting those verbs that showed the clearest differences in patient cardinality preferences between the three types.

The pretest measured patient cardinality preferences for 60 Dutch verbs, 32 of which overlapped with the Hebrew counterparts that were used in Kerem, Friedmann and Winter's pictorial experiment. Specifically, we measured whether there was a

[15] Note that the difference between type 2 and type 3 verbs is not expected to surface when measuring the preference for one-patient situations over two-patient situations (as in Part 1 of Experiment 1). However, the difference between the two types is expected to show when measuring truth-value judgments of reciprocal sentences in situations that contain two-patient situations, i.e. *S6* (in Part 2 of Experiment 1). Therefore, the two groups are distinguished here.

preference between situations with one patient per agent vs. situations with two patients per agent.

Participants

Twenty-one Utrecht University students (19 female, age M = 21) participated for monetary compensation. All participants were native Dutch speakers without dyslexia. Prior to the experiment, all participants signed an informed consent form.

Materials

Sixty different Dutch verbs were tested. These included 16 verbs that were a priori assumed to be of type 1 (neutral verbs like *kennen*, 'know'), 29 verbs that were a priori classified as type 2 (one-patient-preference verbs like *knijpen*, 'pinch'), and 8 verbs that were a priori classified as type 3 (strong one-patient-preference verbs like *bijten*, 'bite'). In addition, we included 7 verbs like *fotograferen* ('photograph') and *ontmoeten* ('meet') (see Table A1 in Appendix A) that were used as filler verbs. These verbs have a typical 'collective' interpretation in which an activity is performed on a collection of patients rather than on individual patients. These verbs were expected to show some preference for two patients over one. They were added to achieve an optimal balance in typicality preferences.

For each verb, we included one experimental pair of schematic representations, reflecting the choice between two instances of that verb – one with one patient and one with two patients. Each schema included three individuals, which were represented by three proper names, and one or more arrows between them, reflecting either one or two actions between the three individuals. An example of an experimental pair appears in the top row of Figure 6.

In addition, for each verb we included five filler pairs in order to control for visual complexity and response bias. The filler items differed from the experimental pair in the number of arrows (Figure 6).

The experimental pairs (1 per verb) and filler pairs (5 per verb) for the 60 verbs resulted in a total of 360 trials. The trials were presented in a pseudo-random order with the restriction that no verb or schematic pair would repeat in two consecutive trials. The position of the different schematic representations (on the left or on the right) and the pointing direction of the arrows (to the left or to the right) were counterbalanced over the trials and the verbs.

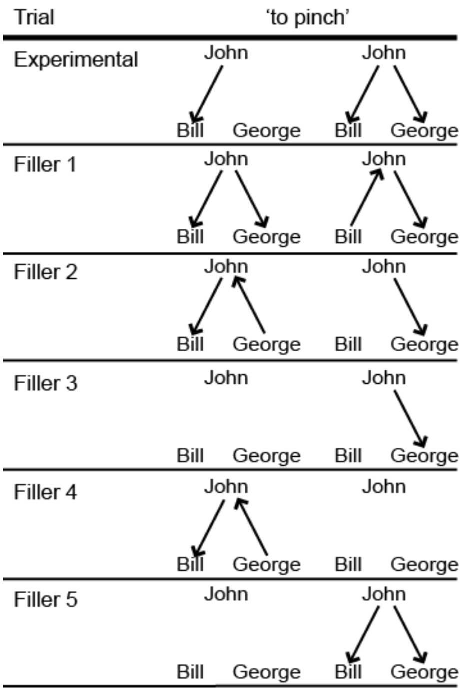

Figure 6: Examples of pairs for *pinch* in the pretest and in Part 1 of Experiment 1 (translated from Dutch)

Procedure

The task was presented in Dutch in a sound-proof booth on a PC using *Presentation* software (Neurobehavioral Systems, Albany, CA). Prior to entering the booth, each participant was instructed verbally about the set-up and on how to interpret the schematic representations. These verbal instructions stressed that each schematic representation should be interpreted as representing a situation at one point in time, rather than multiple situations over a time interval. This was clarified in order to make sure that the verbs are interpreted in the same tense as in Part 2 of Experiment 1 (see footnote 10). Further instructions were given on the PC monitor. After the instructions, each participant completed six practice trials. Subsequently, participants were given the opportunity to ask for further clarifications, followed by six additional practice trials. No verbs that were used in the practice trials were used in the actual pretest. The pretest itself consisted of two blocks of trials. Each trial started with a fixation cross (500ms), followed by the presentation of a verb in the top centre of the screen and, below the verb, a pair of schematic representations: one of the six pairs from Figure 6. Participants were instructed to select the schematic representation that best represented the given verb by pressing the left or right arrow key accordingly, with their dominant hand. The verb and the schematic representations remained visible on the screen for 5000ms, or until the participant responded.

Analysis

We calculated the proportion of reactions to the test items where a schematic representation of a one-patient situation was selected. We performed a correlation analysis on the patient cardinality data on the 32 Hebrew verbs that were tested in Kerem, Friedmann and Winter's pictorial typicality experiment, and the corresponding Dutch verbs that were examined in the current pretest.

Results

The results of the pretest are given in Table A1 of Appendix A. These results about Dutch verbs show a significant one-sided positive correlation with the patient cardinality preferences of the corresponding 32 Hebrew verbs that were tested in Kerem, Friedmann and Winter (r (30) = .39, p (one-tailed) = .014), which indicates that the schematic method yields comparable patient cardinality preferences to the pictorial method of Kerem, Friedmann and Winter (2009). This lends support to our assumption that the two methods measure the same thing, and that Hebrew and Dutch verbs have similar meanings.

The results of the pretest were used to select verbs for Experiment 1. The selection procedure is explained in the Materials section of section 2.4.3.

2.4.3 Experiment 1 (Part 1): Typicality

Part 1 of Experiment 1 is a replication study that measured patient cardinality preferences for a subset of the Dutch verbs that were used in the pretest. This part again measured patient cardinality preferences for the selected subset of verbs. Therefore, we expected to see the same behavior as in the pretest: those verbs that we classified as type 1 verbs were expected to show no preference, those that we classified as type 2 verbs were expected to show a preference for one-patient situations, and those that we classified as type 3 verbs were expected to show the same preference, possibly more substantially.

Participants

Eighteen Utrecht University students (15 female, age M = 21) participated for monetary compensation. All participants were native Dutch speakers without dyslexia. Prior to the experiment, all participants signed an informed consent form.

Materials

Based on the results of the pretest, we selected 18 verbs that were to be tested regarding their patient cardinality preferences (see Table A2 in Appendix A). The selection process went as follows. Firstly, to avoid confounds we ruled out verbs that could not be expressed in one word (e.g. *luisteren naar* ‘listen to’, *muziek maken* ‘make music’).[16] Secondly, we selected six verbs from each of the three classes of verbs. From the type 1 verbs (neutral), we selected the six verbs that proved to be the most neutral in their preference between a one-patient situation and a two-patient situation. From the type 2 verbs (one-patient-preference) as well as the type 3 verbs (strong one-patient-preference), we selected the six verbs from each class that showed the highest preference for a one-patient situation.

The 18 verbs selected for Experiment 2 are the Dutch correlates to the following verbs[17]:

Type 1 (neutral):	*envy, know, understand, admire, miss, hate*
Type 2 (one-patient-preference):	*pinch, hit, caress, stab, shoot, grab*
Type 3 (strong one-patient-preference):	*kiss, dress, kick, lash out, bite, lick*

Similarly to the pretest, for each of the selected 18 verbs we used one experimental pair and five filler pairs to control for visual complexity and frequency of arrow combinations (Figure 6). This resulted in a total of 108 trial items, which were presented in a pseudo-random order with the restriction that no verb or schematic pair would repeat in two consecutive trials. The position of the different schematic representations (on the left or on the right) and the pointing direction of the arrows (to the left or to the right) were counterbalanced over the trials and the verbs.

Procedure

The procedure of the task in Part 1 was identical to the procedure of the pretest.

Analysis

We calculated the proportion of reactions to the test items where a schematic representation of a one-patient situation was selected. We then performed a correlation analysis on this patient cardinality data and the patient cardinality data from the same verbs in the pretest, in order to verify that the results were comparable. Next, we performed an ANOVA with Verb as the random variable and Verb type as the between-item variable with three levels (1, 2 and 3). Post-hoc Bonferroni corrected pairwise comparisons were performed to analyze the differences between the verb types. A

[16] These verbs were only included in the pretest to test the correlation with the Hebrew results from Kerem, Friedmann and Winter (2009).

[17] Notice that type 1 verbs are all stative verbs while type 2 and type 3 verbs are action verbs. This is simply the way that the verbs grouped together after we determined the selection criteria for each group. It would be interesting to investigate why they group together this way, i.e. why it is that stative verbs and action verbs show different typicality behaviour. For the current work, however, it does not affect our argument since we are interested in the overall correlation between typicality and interpretation – regardless of a verb’s type (see also footnote 10).

repeated measures ANOVA with Participants as the random variable and Verb type as the within-subject variable, again with three levels, gave similar results as the item analysis, therefore we only report the first analysis.

Results

We found a significant correlation (r (16) =.58, p (one-tailed) = .006) between the results for the 18 verbs in Part 1 of Experiment 1 and those same 18 verbs in the pretest. That means that as expected, we replicated the pretest for the 18 selected verbs.

For the 18 verbs in Part 1 of Experiment 1, the percentage preference for one-patient situation ranged from 33% to 100%. The ANOVA revealed a significant main effect of Verb type (F (2,15) = 27.40, p < .001). As expected, the results show a strong preference for one-patient situations with both type 2 verbs (M = .83, SD = .10) and type 3 verbs (M = .89, SD = .13). Pairwise comparisons showed that there was no significant difference between these preferences (t(10) = 0.91, p = 1.00). The preference for one-patient situations with type 1 verbs (M = .43, SD = .13) was significantly lower compared to type 2 verbs (t(10) = 5.90, p < .001) and type 3 verbs (t(10) = 6.82, p < .001). See Figure 7, and the further details in Table A2 in Appendix A.

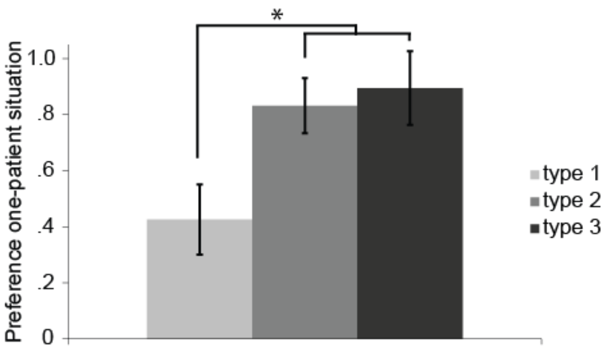

Figure 7: Experiment 1 Part 1 – analysis of preference for a one-patient situation with three types of verbs.

2.4.4 Experiment 1 (Part 2): Reciprocal interpretation

Part 2 of Experiment 1 measured acceptability of Dutch reciprocal sentences in different situations. The verbs that were tested were the same 18 verbs that were used in Part 1, again using schematic representations. We used a truth-value judgment task to measure to what extent a given reciprocal sentence is accepted in each of the situations *S6* and *S3*. These two situations were expected to be core situations for the different sentences that we tested: *S6* for sentences with verbs that show no patient cardinality preferences, and *S3* for sentences with verbs that show a preference for one-patient situations. Accordingly, we expected to see different acceptance rate patterns for *S6* and *S3* with the three types of verbs. Hence, in contrast to Kerem, Friedmann and Winter's study (2009), where a forced-choice task was used, Experiment 1 measured acceptance rates of reciprocal sentences in different situations in a truth-value judgment task. This task allowed us to test the predictions of the MTH.

Participants

A total of 25 Utrecht University students participated for monetary compensation (24 female, age M = 23). All participants were native Dutch speakers without dyslexia. Prior to the experiment all participants signed an informed consent.

Materials

The same 18 verbs from Part 1 of Experiment 1 were used, but now in Dutch reciprocal sentences of the form *A, B and C P each other* (where *A*, *B* and *C* are proper names and *P* is a verb). The resulting 18 sentences were tested for their interpretation in a truth-value judgment task.

For each sentence, we included two experimental trials – schemas reflecting *S6* and *S3*. Those situations illustrate relations between three individuals. In situation *S6,* each individual acts on both other individuals. In situation *S3,* each individual acts on exactly one other individual and is acted on by exactly one other individual. Each schema included three individuals, which were represented by three proper names, and either six arrows (in *S6*) or three arrows (in *S3*) between them. Examples of these experimental trials are in the top two rows of Figure 8.

In addition, for each verb we included one control trial and two filler trials (Figure 8). The filler trials were used to control for visual complexity and frequency of no-responses. The control trials illustrated a situation in which only two individuals act on another individual (*S2*). We added this control item because it was expected to be a typical instance for all verbs, but not the core situation for reciprocal sentences containing them. Similarly, for sentences with verbs of type 1, situation *S3* is expected to be typical but not the core situation. By adding the control trial *S2*, we aimed to also have typical but non-core situations for sentences with verbs of types 2 and 3. This allowed us to better test the predictions of the proposed MTH since this principle also makes predictions about acceptability rates for non-core situations.

The experimental trials (2 per verb), control trial (1 per verb) and filler trials (2 per verb) for all 18 sentences resulted in a total of 90 trials. The trials were presented in a pseudo-random order with the restriction that no verb or schematic representation would repeat in two consecutive items. The pointing direction of the arrows (to the left or to the right and up or down) was counterbalanced over the verbs.

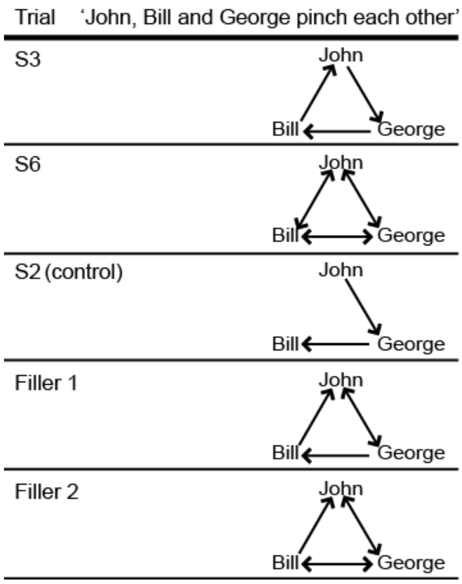

Figure 8: Examples of trials for *John, Bill and George pinch each other* in Part 2 of Experiment 1 (translated from Dutch)

Procedure

Part 2 of Experiment 1 consisted of two blocks of trials. As in Part 1, the task was presented in Dutch in a sound-proof booth on a PC with *Presentation* software (Neurobehavioral Systems, Albany, CA). The instructions and practice trials prior to the experiment also resembled those of Part 1.

Each trial started with a fixation cross (500ms), followed by the presentation of a reciprocal sentence (e.g. *John, Bill en George knijpen elkaar*, 'John, Bill and George pinch each other') at the top of the screen. After 2000ms, a schematic representation was added to the screen, below the sentence. Participants were instructed to indicate whether the situation that was presented schematically is a possible depiction of the sentence or not, by pressing a green or red button accordingly (right and left arrow key respectively, marked with a sticker), with their dominant hand. The sentence and the schematic representation remained visible on the screen until the participant responded, or for 10000ms if there was no response.

Analysis

We calculated the proportions of positive responses to the control trials (*S2*) and the experimental trials (*S6* and *S3*), which reflects the acceptability of a given schematic representation of a reciprocal situation as a possible depiction of the given sentence. Further statistical analysis focused on the experimental trials (*S6* and *S3*). We performed a mixed ANOVA with Verb as the random variable, Verb type as the between-item variable (with three levels: 1, 2 and 3) and Trial type as within-item variable (with two levels: *S6* and *S3*). Post-hoc Bonferroni corrected pairwise comparisons were performed to analyze the differences between the verb types. A repeated measures ANOVA with Participants as the random variable and Verb type (1, 2 and 3) and Trial type (*S6* and *S3*) as the within-subject variables, gave similar results as the item analysis, therefore we only report the first analysis.

Results

Control trial S2: The acceptance of reciprocal sentences in *S2* was very low for sentences with all verb types: reciprocal sentences with type 1 verbs (M = .06, SD = .04), reciprocal sentences with type 2 verbs (M = .13, SD = .03), and reciprocal sentences with type 3 verbs (M = .12, SD = .01). For details see Table A2 in Appendix A.

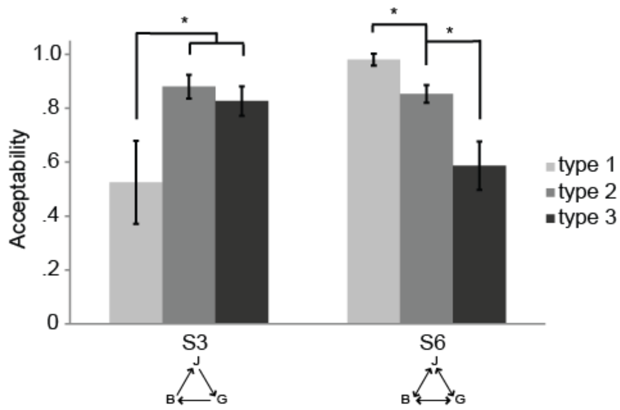

Figure 9: Experiment 1 Part 2 – item analysis of acceptance rate for sentences in *S6* and *S3* with 3 types of verbs.

Further analysis of the results of Part 2 of Experiment 1 focused on the two experimental trial types that measured acceptability of reciprocal sentences in *S6* and *S3*. There was a significant main effect of Trial type (F (1,15) = 4.63, p = .048) and Verb type (F (2,15) = 16.01, p < .001), as well as a significant interaction between Trial type*Verb type (F (2,15) = 49.67, p < .001). This interaction was further analyzed with pairwise comparisons.

Experimental trial S6: Regarding reciprocal sentences in *S6*, acceptability percentages ranged from 44% to 100% per sentence. There were significant differences between all three types of tested sentences (all p's ≤ .005) (see Figure 9 and Table A2 in Appendix A). The highest acceptability of sentences in *S6* was found for those sentences containing type 1 verbs (M = .98, SD = .02), followed by those with type 2 verbs (M = .85, SD = .03) and those with type 3 verbs (M = .59, SD = .09).

Experimental trial S3: Regarding acceptance of reciprocal sentences in *S3* (which ranged from 36% to 96% per sentence), the pattern was different (see Figure 9 and Table A2 in Appendix A). There was no significant difference between reciprocal sentences with type 2 verbs (M = .88, SD = .04) and reciprocal sentences with type 3 verbs (M = .83, SD = .05, t (10) = .95, p = 1.00). Both types of sentences showed high levels of acceptability in *S3*. The acceptability of reciprocal sentences with type 1 verbs in *S3* (M = .53, SD = .15) was significantly lower compared to sentences with type 2 verbs (t (10) = 6.34, p < .001) and with type 3 verbs (t (10) = 5.39, p < .001).

2.4.5 Correlation between results of Part 1 and Part 2

As expected, the results from Part 1 showed a considerable variability in the patient cardinality preferences of different verbs (M = .72, SD = .24, Table A2 in Appendix A). The MTH predicts that the preference for one-patient situations (as measured in Part 1) correlates positively with the acceptability of reciprocal sentences in *S3* (as measured in Part 2), explaining the interpretations of reciprocal sentences with type 2 verbs. This prediction was borne out: we found a significant one-sided positive correlation (r (16) = .76, p < .001, see Figure 10).

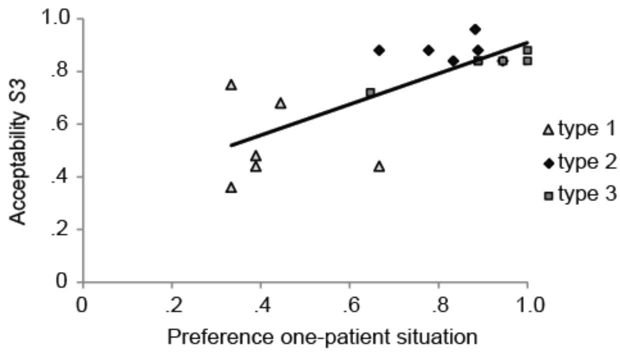

Figure 10: Relation between preferences for one-patient situations in Part 1 of Experiment 1, and acceptance of reciprocal sentences in *S3* in Part 2 of Experiment 1.

2.4.6 Discussion

The main aim of Experiment 1 was to test the predictions of the proposed Maximal Typicality Hypothesis: whether the core situation described by a reciprocal sentence is maximal among those situations that are most typical for the verb concept in the reciprocal's scope. To that end, Part 1 measured typicality effects for 18 different verb concepts, which were selected based on their patient cardinality preferences as measured in a pretest. Part 2 measured whether reciprocal sentences containing those 18 verbs were accepted in *S6* and/or *S3* (i.e. the two experimental trials). The test for the MTH was in the observed differences between verb types, and in the more fine-grained correlation analysis between the two parts of Experiment 1.

Main results: the core situation

In Part 1 of Experiment 1, we found that different verbs show different patient cardinality preferences. As expected, those verbs that we a priori classified as type 1 (neutral) showed no clear preference for an instance with one patient vs. two patients. By contrast, those verbs that we called type 2 (one-patient-preference) and type 3 (strong one-patient-preference) showed a clear preference for instances with one patient over instances with two patients. From these patient cardinality preferences, we extrapolate typicality preferences for reciprocal situations. For type 1 verbs, we extrapolate that there is no difference in typicality between reciprocal situations that differ in terms of patient cardinality, while for verbs of type 2 and 3 we extrapolate that reciprocal situations with one patient per agent are more typical than those with two patients per agent. If, as extrapolated, all situations for type 1 verbs are of equal typicality, then the MTH predicts that the core situation for a reciprocal sentence with a type 1 verb is *S*6: the maximal situation among all situations. Thus, for reciprocal sentences with type 1 verbs the MTH expects the acceptability in *S6* in Part 2 of the experiment to be high. For sentences with type 2 and type 3 verbs, the MTH predicts the core situation to be *S3*: the maximal among those situations that respect the verbs' typicality preference for one-patient situations. Thus, the MTH predicts the acceptability of these sentences to be high in *S3*.

Part 2 supports these predictions of the MTH. Reciprocal sentences with type 1 verbs are very often accepted in *S6* (98% of the time), significantly more so than sentences with type 2 and 3 verbs. Our findings also indicate that sentences with type 3 verbs are highly acceptable in *S3* (83%), significantly more so than sentences with type 1 verbs. These findings are expected by the MTH. The same results can also be explained by the SMH or MIH, if it is assumed that type 3 verbs like *bite* are judged as physically impossible in situations with two patients per agent (an assumption that we did not test directly in our experiments). However, our findings on type 2 verbs are substantial support for the MTH over previous hypotheses. These findings indicate that reciprocal sentences with such verbs are highly acceptable in *S3* (88%), significantly more so than sentences with type 1 verbs. This is as expected by the MTH, since it selects *S3* as the core situation for sentences with type 2 verbs. The SMH and MIH, however, do not explain the acceptability of sentences with type 2 verbs in *S3*, since they predict only one meaning for such sentences – Strong Reciprocity, which does not hold in *S3*.

The success of the MTH in predicting the acceptability pattern for reciprocal sentences in *S3* is further supported by the significant correlation that was found between this acceptability and preferences for one-patient situations as measured in Part 1. In this way, the MTH not only explains the overall acceptability of reciprocal sentences with type 2 verbs in *S3*, but gives a fine-grained prediction for the interpretation of other reciprocal sentences in *S3* situations.

Additional results: non-core situations

So far, we have only discussed the acceptability of reciprocal sentences in the core situation. That is, the high acceptability of sentences with type 1 verbs in *S6*, and the high acceptability of sentences with type 2 and type 3 verbs in *S3*. In addition to that, however, our data provide information about the acceptability of these sentences in non-core situations.

Firstly, one important aspect of our results is the acceptability of sentences with type 2 and type 3 verbs in *S6*. For both kinds of sentences, the MTH predicts the same core situation: *S3*. However, as we might expect, the acceptability of reciprocal sentences with type 3 verbs (e.g. *bite*) in *S6* is significantly lower than the acceptability of sentences with type 2 verbs (e.g. *pinch*) (t (10) = -4.71, p < .001). Sentences with type 2 verbs were in fact equally acceptable in *S6* and *S3* (t (10) = 0.05, p = .605). Potential differences between type 2 verbs and type 3 verbs with respect to two-patient situations like *S6* were not measured in the preference tasks of Part 1. Thus, the difference in interpretations between sentences with type 2 verbs and type 3 verbs calls for a separate explanation. We hypothesize that the high preference for one-patient situations with such verbs (in Part 1) has a different source with type 3 verbs than with type 2 verbs. In the case of type 3 verbs, we believe that for many participants the one-patient preference reflects not merely a choice between two possible situations, but a preference of a possible situation (with one patient) over an impossible, or inconceivable, situation (with two patients). For type 2 verbs, one-patient situations are uniformly preferred over two-patient situations, but the latter are likely to be accepted as instances of the verb concept, as witnessed by the high acceptability rates of sentences with such verbs in *S6*. Such a difference between type 2 and type 3 verbs could not be measured in the forced-choice task of Part 1, but it would immediately affect acceptability judgments of reciprocal sentences in *S6* in Part 2. If a participant thinks that a two-patient situation is not possible for type 3 verbs like *bite*, then she will judge *S6* as strictly speaking impossible for a reciprocal sentence with such verbs. For type 2 verbs, two-patient situations put less strain on the imagination of the participants. Accordingly, *S6* situations are usually accepted for reciprocal sentences with type 2 verbs.[18] This pattern with *S6* situations and type 2 and 3 verbs does not come as a surprise: it fully agrees with our *a priori* classification of type 2 and type 3 verbs, which was based on mere introspection, and it is consistent with the results of Kerem, Friedmann and Winter (2009) on type 2 verbs, and the intuitive SMH-based

[18]A similar point could be seen in in the forced-choice task in Kerem, Friedmann and Winter (2009), where *S6* was often preferred to *S3* for type 2 verbs, despite the high typicality of one-patient situations.

treatment of type 3 verbs. Therefore, we believe that simple experimental measures can distinguish type 2 verbs from type 3 verbs: e.g. asking participants to mark possible situations, rather than to choose between them. Running such an experiment would be unproblematic, if relevant for further research.

Another question on non-core situations concerns the status of *S3* for sentences with type 1 verbs, which is properly contained in their core situation *S6*. As expected, reciprocal sentences containing type 1 verbs are fully acceptable in *S6*. The SMH and MIH account for this fact using the Strong Reciprocity operator, which expects these reciprocal sentences to be downright unacceptable in all situations that are properly contained by *S6*. The acceptability rates of sentences with type 1 verbs in *S3* in our experiment (M = .53, SD = .15) go against this prediction of the SMH and MIH, especially when compared to the downright unacceptability of the same sentences in *S2*. This problem for the SMH and MIH arises because the hypotheses make standard binary (true/false) predictions about reciprocal sentences. Accordingly, the SMH and MIH expect decisive judgments on sentences in all situations. Unlike those hypotheses, the MTH does not use such absolute terms for acceptability of reciprocal sentences in situations that are properly contained in the core situation. The MTH only expects this acceptability to be lower than the acceptability in the core situation. Thus, sentences with type 1 verbs are expected to show decreasing acceptability in *S2* and *S3* situations compared to *S6*. Similarly, sentences with type 2 and 3 verbs are expected to show lower acceptability in *S2* than in *S3*.

Hence, our acceptability results on *S2* and *S3* are consistent with these predictions. Notwithstanding, the MTH does not give a quantitative account of the large gap between the judgments on *S2* and *S3* with type 1 verbs. Thus, the question for the MTH is what explains the relatively high acceptability of reciprocal sentences with type 1 verbs in *S3* compared to *S2*. We speculate that there are at least two possibilities. Under the first possibility, the difference is the result of something independent of *S2* and *S3* in particular. It might be that the threshold for category membership for the particular verbs that we tested is between *S2* and *S3*. This means that the threshold might be somewhere else for different verbs: for the concept *stacked on top of* the threshold might be below *S2*. Under the second possibility, *S3* situations have a special status with reciprocal sentences: even though they are not maximal situations, they still guarantee that each agent is acting on some patient, and that every patient is acted upon by some agent. This is the requirement imposed by the logical operator that Langendoen (1978), and more subsequent work, assume to be the (unique) logical operator responsible for interpreting reciprocal expressions: *Weak Reciprocity* (see subsection 2.2.1). Pragmatic theories using the weak reciprocity operator as the default reciprocal meaning do not account for our data, since similarly to Dalrymple et al.'s SMH and Sabato and Winter's MIH, these theories assume bi-valued judgments on all situations. At the same time, we believe that satisfaction of Weak Reciprocity can lead to moderate acceptance of *S3*, even when this situation is not among the most typical ones, hence not the core situation. Experimental exploration of this hypothesis about the privileged status of *S3* as a reciprocal situation is deferred to further research.

2.5 Conclusion

This chapter started with a multitude of operators that were proposed as meanings of the reciprocal expression *each other*, each making a different logical contribution to the reciprocal sentence as a whole. The many different interpretations of reciprocal sentences have raised the question which principles affect the availability of those interpretations, and in particular what the role of context is. While the SMH and the MIH both take into account the verb in the scope of the reciprocal expression to some degree, neither account makes the correct predictions for reciprocal sentences with type 2 predicates. Our experiment shows that these sentences are accepted in *S3* situations despite the fact that they are also accepted in *S6* situations. Following Kerem, Friedmann and Winter, we proposed the Maximal Typicality Hypothesis, which draws on the SMH but refines the notion of context that is relevant for determining reciprocal interpretation. The MTH is a principle that specifies the core situation for reciprocal sentences based on typicality effects for the verb concept in the reciprocal's scope. Specifically, it predicts the core situation to be the situation that is maximal among the most typical ones for the verb concept. This hypothesis correctly predicts the acceptability pattern for sentences with type 2 verbs. More generally, the results reveal a new semantic perspective on typicality effects: in addition to their well-studied prominence in categorization, typicality also interacts with truth-conditional aspects of sentential meaning.

Chapter 3
Plural predicate conjunction & typicality effects

Parts of this chapter also appeared in Poortman (2014) and Poortman (to appear).

3.1 Introduction

The current chapter studies the interpretation of sentences containing conjoined natural language predicates (nouns, adjectives and verbs). Under the standard semantic view, the denotation of a predicate is analyzed as a set, and conjunction of predicates is analyzed as set intersection (e.g. Gazdar, 1980; Partee & Rooth, 1983; Keenan & Faltz, 1985). This is often referred to as 'Boolean conjunction', since set intersection is an instance of the 'meet operation' in Boolean algebras. Examples of simple singular sentences with conjoined adjectival, nominal and verbal predicates are in (1)-(3), respectively.

(1) John is tall and slim
(2) John is a good father and an excellent husband
(3) John is singing and dancing

Sentence (1) is generally considered to be true if and only if John is in the intersection of the two sets that are denoted by the conjoined adjectives. In other words, it is true only in situations in which John is tall and John is slim. Similarly, (2) is true iff John is a good father and John is an excellent husband, and (3) is true iff John is singing and John is dancing.

The focus of this chapter is on plural sentences with predicate conjunction. These conjunctions are a case for which the intersection analysis sometimes fails. For example, let us consider the following sentences (4) and (5).

(4) The men are sitting and reading
(5) The men are waving and smiling

Plural sentences with conjunctive predicates as in (4) and (5) are similar to (1)-(3), the main difference being in the number of individuals that is predicated over. Previous works have therefore applied the same Boolean analysis of conjunction to (4) and (5) as for (1)-(3), and combined it with a distributivity operator which assumes universal quantification (Link, 1983; 1987; Partee & Rooth, 1983; Keenan & Faltz, 1985). That means that the properties described by the verbs in each sentence are distributed over all individuals that make up the plural subject. Accordingly, the operator shifts a VP like *waving and smiling* which holds of an individual (like *John*) into a VP that can hold of a plural individual (like *the men*), iff the VP holds of each atomic part of the plural individual (i.e. each man). Thus, sentences (4) and (5) are expected to be true if and only if every man is in the intersection of the two sets that are denoted by the conjoined verbs. Hence, (4) is expected to be true only in situations in which each man is sitting and each man is reading, and (5) is expected to be true only in situations in which each man is waving and each man is smiling. While this seems straightforward, there are reasons to doubt that the Boolean analysis for singulars can be extended to plurals this way. Consider sentence (6), which is minimally different from (4).

(6) The men are sitting and standing

Analogously to (4), sentence (6) might be expected to be true only in situations in which each man is sitting and each man is standing. However, such situations contradict our world knowledge: a man cannot simultaneously sit and stand. Sentence (6) is most naturally interpreted such that a subset of the men is sitting and the rest of the men are standing. I henceforth refer to the situations that such an analysis describes as 'split situations' (Heycock & Zamparelli, 2005). Since this analysis is clearly not consistent with Boolean conjunction, cases such as (6) have been used to argue for the existence of non-Boolean conjunction (Link, 1983; 1984; Krifka, 1990). In this chapter, I will argue that the simple examples that earlier works have considered are part of a special class, which represent only a part of the data on the phenomenon of predicate conjunction: those that lead to a clear Boolean analysis and those that lead to a clear non-Boolean analysis. Examples such as (4) and (5) contain 'compatible predicates': according to our world knowledge, it is possible to apply a Boolean analysis in which both properties are distributed via universal quantification. Such examples have been used to argue in favor of Boolean conjunction (e.g. Winter, 2001a). By contrast, examples such as (6) contain 'incompatible predicates': applying a Boolean analysis leads to an interpretation that contradicts our world knowledge. Hence, such examples have been used to argue for non-Boolean conjunction (e.g. Krifka, 1990). I will show that this sharp contrast between compatibility and incompatibility of predicates fails to capture all the facts on sentence interpretation. Instead, testing actual speakers' behavior is required in order to systematically measure the correct notion of compatibility, which turns out to be graded. Consider sentence (7).

(7) The men are sitting and cooking

Sentence (7) contains predicates that are not strictly incompatible, but also not naturally compatible. It is possible for each individual man to sit and cook simultaneously, but it is not a very likely situation. As we will see, sentence (7) is accepted in a split situation, but crucially to a lesser extent than sentence (6).

Besides claiming that the examples that have so far been studied (e.g. (4)-(6)) do not represent the classes of data that are relevant for a theory of conjunction, I will show that even those examples are not as clear-cut as was claimed based on introspection. My experimental measures reveal that sentences with presumed compatible verbal predicates as in (4) and (5) are acceptable in split situations on average 54% of the time. This means that even seemingly simple cases cannot straightforwardly distinguish between a Boolean and a non-Boolean analysis of conjunction. Moreover, under either hypothesis, it needs to be explained why (6) is accepted in a split situation significantly more often than (4) and (5).

The main question throughout the chapter will be what explains the variability in logical interpretations between structurally similar plural sentences with conjoined predicates, specifically adjectival and verbal predicates. In order to answer that question, I use experimental measures that allow studying the relevant contextual-lexical effects and hence systematically evaluate previous proposals.

Krifka (1990) proposes to extend the generally accepted non-intersective conjunction of NPs like *John and Mary* to conjunction of predicates like *sitting and reading* or *sitting and standing*. He argues that the meaning of conjunction is generally weak, and strengthened only when pragmatics allows it. However, while this may be a proper analysis for sentences like (6), I provide substantial experimental evidence that shows that such a weak account does not work without further assumptions. Firstly, sentences like (4) and (5) are often rejected in weak, split situations. And secondly, sentences like (7) are more often rejected in such split situations than sentences like (6). Without further assumptions, Krifka (1990) does not account for these differences. Later, Winter (2001a) claims that on top of Krifka's descriptive proposal of non-intersective conjunction, we also need a principle that determines when which analysis is allowed, thus when the different interpretations are actually derived. Winter proposes that a maximality principle like the SMH – which was discussed in the context of reciprocal sentences in Chapter 2 - is a suitable candidate. He adapts the SMH so that it selects the strongest possible sentence meaning for plural sentences with predicate conjunction just like it does for reciprocal sentences. In a nutshell, non-intersective interpretations are only available when strong, intersective interpretations are ruled out by the nature of the conjoined predicates (e.g. for a combination like *sitting and standing*). The experimental data in this chapter show that such an account also fails to fully account for the facts. If Winter's extended version of the SMH (eSMH) is correct, then sentences like (4), (5), and even (7) – whose conjoined predicates do not rule out an intersective interpretation – should only receive such a strong, intersective interpretation. My data clearly show that this is not the case: such sentences are accepted in a split situation ranging from 50% to nearly 80% of the time.

In sum, there are two problems for previous proposals. Firstly, both Krifka's and Winter's proposals are not fine-grained enough because they cannot distinguish between sentences with strictly incompatible verb pairs like (6) and sentences with very unnatural, though not strictly incompatible pairs like (7). Secondly, Krifka and Winter do not make the correct predictions for simple cases with naturally compatible pairs like (4) and (5), that are accepted in split situations about 50% of the time: these accounts expect those sentences to be uniformly judged as true/false, respectively. More generally, this acceptability pattern reveals that truth conditions cannot be stated in absolute terms, independent of a speaker's knowledge of the world. To solve these problems for previous proposals, I propose that truth conditions should be stated relative to a speaker's conceptual knowledge. Understanding the acceptability pattern for the sentences I have discussed calls for a more systematic investigation of the conceptual information that is tied to the predicates in the sentences than merely in terms of compatibility vs. incompatibility. This is similar to the findings on reciprocal sentences in Chapter 2. Consequently, I propose that the Maximal Typicality Hypothesis (MTH) of Chapter 2 can be extended to account for the experimental data on predicate conjunction. Again, the MTH specifies the core situation for a sentence based on typicality features of the verb concepts in that sentence.

The chapter is structured as follows. Section 3.2 reviews some background on conjunction. It starts by briefly describing some of the problems for an intersection analysis of conjunction that have been researched extensively, and then zooms in on the specific problematic area of plural predicate conjunction which has been investigated less extensively. Section 3.3 presents problems for current proposals on predicate conjunction by summarizing the main new data. Next, section 3.4 argues that the Maximal Typicality Hypothesis that was proposed as an account of plural reciprocal sentences also makes the right predictions for plural conjunctive sentences. Sections 3.5 and 3.6 describe new experimental work that uses original measures that were specifically constructed to test the eSMH and MTH in this distinct domain: experiments on adjectival predicate conjunction and verbal predicate conjunction, respectively. Following the experimental work, 3.7 describes some remaining discussion points. Finally, section 3.8 concludes this chapter.

3.2 Some background on conjunction

Subsection 3.2.1 first describes some of the issues for an intersection analysis of conjunction that have been researched extensively. These include the conjunction of two argument noun phrases like *John and Mary*, and the conjunctions of singular nominals as in *this soldier and sailor*, as well as plural nominals as in *my friends and colleagues*. Next, subsection 3.2.2 zooms in on the particular problematic construction that is in the focus of this chapter: the conjunction of adjectival or verbal predicates within plural sentences, e.g. *big and small* or *sitting and cooking*.

3.2.1 Well-known problems for an intersection analysis of conjunction

Conjunction of two argument noun phrases and the collective/distributive distinction
One classic issue in the area of conjunction is the conjunction of noun phrases and the difference between collective and distributive readings of sentences. Consider sentences (8) and (9).

(8) John and Mary danced
(9) John and Mary are asleep

Conjunction of two argument noun phrases like *John and Mary* cannot be analyzed intersectively in the same way as the conjunction *tall and thin* in (1) was analyzed, for example. The conjunction *John and Mary* does not denote the intersection of the property of being Mary and of being John. However, an intersection analysis can be saved for some cases once we treat noun phrases as generalized quantifiers (e.g. Barwise & Cooper, 1981). Under that assumption, a noun phrase like *John* actually denotes the set of properties that John has. As a result, *and* in (8) and (9) can still be analyzed intersectively: it intersects the set of properties that are possessed by John with the set of properties that are possessed by Mary. In order for (8) to be true then, the property *dance* should be in that intersection, and similarly for the predicate *is asleep* in (9). That means that (8) is true in situations in which John is dancing and Mary is dancing, and (9) is true in situations in which John is asleep and Mary is asleep. A problem for the intersection analysis arises once we also consider sentences like (10) and (11).

(10) John and Mary met
(11) John and Mary are a happy couple

Despite their similar form containing a subject noun phrase with conjoined names and a verb phrase, the logical properties of (10) and (11) are different from the logical properties of (8) and (9). This is immediately apparent once we consider the following inference patterns.

(8) John and Mary danced $\leftrightarrow$ John danced and Mary danced
(9) John and Mary are asleep $\leftrightarrow$ John is asleep and Mary is asleep

(10) John and Mary met $\not\leftrightarrow$ *John met and Mary met
(11) John and Mary are a happy couple $\not\leftrightarrow$ *John is a happy couple and Mary is a happy couple

Sentences (8) and (9) are logically equivalent to their counterparts with sentential conjunction, while (10) and (11) are not. In fact, their sentential counterparts are not even grammatical. The predicates *meet* and *are a happy couple* do not denote properties that can be possessed by either John or Mary individually. Instead, in sentences (10) and (11) some sort of plurality is formed by two entities and a property is predicated of that collective individual instead of distributed over both individuals. This is known as the collective/distributive puzzle for sentences with conjoined noun phrases.

For such cases of collective predication as (10) and (11), different solutions have been offered in the literature. Some authors have taken those cases as evidence for non-Boolean conjunction, which transforms individuals into a plural individual (their 'sum' or 'collection') in order for a collective predicate like *meet* to apply to it (see Massey, 1976; Link, 1983; Hoeksema, 1983; Krifka, 1990; Heycock & Zamparelli, 2005). Others have adapted an intersection analysis of conjunction so that it can apply to both types of examples (see Winter, 2001b; Champollion, 2015). For the purposes of this dissertation, it is not necessary to distinguish between these solutions, as will become clear from section 3.2.2 and onwards.

Conjunction of singular nominals

Another case where an intersection analysis sometimes fails, is noun phrase conjunction inside a determiner phrase (Heycock & Zamparelli, 2005). Consider first sentences (12) and (13), taken from Heycock & Zamparelli (2005).

(12) My friend and colleague is writing a paper
(13) That liar and cheat is not to be trusted

In sentences (12) and (13), the determiner phrases *my friend and colleague* and *that liar and cheat* each refer to a single individual that has two properties. This is in line with an intersection analysis of conjunction, again assuming that noun phrases can denote predicates: *and* in (12) intersects the set of friends with the set of colleagues, and *my friend and colleague* refers to the individual that is in that intersection. Unexpectedly however, this analysis does not apply to (14) and (15).

(14) This soldier and sailor are inseparable
(15) This man and woman are in love

The determiner phrase in (14) does not refer to a single individual that has the property of being a soldier and a sailor. Instead, *this soldier and sailor* refers to a pair of individuals. This is easily visible from the plural agreement and the collective meaning of the verb *to be inseparable*. The same reasoning holds for *this man and woman* in (15), as is visible from the plural agreement, the collective meaning of *to be in love*, and in this particular example also the lexical information regarding the impossibility of *man and woman* to refer to a single individual. Each predicate in the conjunction applies to one of the individuals in the pair. Heycock & Zamparelli call this the 'split interpretation', in contrast to the 'joint interpretation' of (12)-(13).[19] Importantly, due to verbal agreement, the determiner phrases in (14) and (15) unambiguously lead to a split interpretation while the determiner phrases in (12) and (13) unambiguously lead to a joint interpretation.

Every-constructions as in (16) and (17) can also lead to split interpretations, despite the singular agreement (Bergmann, 1982; Winter, 1998). Notably, such sentences are ambiguous between a split and a joint reading when they occur out of context.

(16) Every linguist and philosopher knows the Gödel Theorem
(17) Every officer and gentleman attended the event

Sentence (16), for example, can either claim that every individual who is both a linguist and a philosopher knows the theorem (the joint reading) or that every linguist and every philosopher knows the theorem (the split reading).

Conjunction of plural nominals

The apparently non-intersective, split behavior of conjoined singular noun phrases within a determiner phrase also extends to plural noun phrases (Heycock & Zamparelli, 2005; Champollion, 2015). A split reading can appear unambiguously due to incompatible properties such as in *men and women* or *parents and grandparents* in (18)-(19). Or it can appear ambiguously alongside an intersective interpretation, as in (20)-(21). Note that for plural conjunctions as in (20) and (21), there is no way to disambiguate between the two readings without context. Unlike sentences with singular nominals as in (14) and (15), verbal agreement in the plural sentences does not distinguish between split and joint interpretations. In addition, lexical information regarding the incompatibility of nominals also does not distinguish between the two possibilities here: it is perfectly possible for one individual to be both a sailor and a soldier, or both a linguist and a philosopher.

---

[19] With singular nominal predicates, such split interpretations do not appear in all languages. While they appear for example in English, Dutch and Finnish, they do not appear in German, French, Italian and Spanish (Dowty, 1988; King & Dalrymple, 2004; Heycock & Zamparelli, 2005).

(18) Ten men and women got married today in San Pietro
(19) My parents and grandparents gathered for the wedding
(20) These sailors and soldiers are often together
(21) These linguists and philosophers all came to the party

The main thing that examples (18)-(21) teach us is that there exist non-intersective interpretations of nominal conjunction in plural sentences, just like nominal conjunction in singular sentences as in (14)-(17). I will come back to this in section 3.7.3.

3.2.2 Conjunction of plural adjectival or verbal predicates

The current subsection zooms in on the specific puzzle that is in the focus of this chapter. That puzzle concerns the interpretation of plural sentences with conjoined adjectival or verbal predicates. Unlike the challenges with conjoined nominals that were described in subsection 3.2.1, the current challenge has not yet been researched extensively. Throughout the chapter, the main claim is that detailed lexical information on content words is required in order to explain the observed acceptability patterns, similar to the reciprocal sentences in Chapter 2. Mechanisms that were previously proposed for examples in subsection 3.2.1 do not suffice. While some researchers have assumed a general collective theory of conjunction (e.g. Krifka, 1990; Heycock & Zamparelli, 2005) and others have argued for a theory of conjunction that is strictly intersective (Winter, 1998; Champollion, 2015), the remainder of this chapter aims to convey that none of those proposals can do without taking into account the effects of content words on truth-conditions. I do this without assuming any specific theoretical account. In principle, the proposal that I put forward – the MTH – can be implemented in all sorts of semantic formalisms (e.g. as a weakening mechanism or as a strengthening mechanism). Choosing between those is not the goal of my work.

Previous proposals dealing with conjunction of adjectival and verbal predicates are Krifka's (1990) general proposal of non-Boolean conjunction and Winter's (2001a) extended Strongest Meaning Hypothesis, which aims to save a Boolean analysis of conjunction. Krifka (1990) argues against a Boolean analysis of predicate conjunction. As evidence in favor of this proposal, he provides examples of non-Boolean conjunction of adjectives and verbs as in (22) and (23) (also noticed by Lasersohn (1988)).

(22) The flag is green and white
(23) The dogs and the roosters barked and crowed all night

Intuitively, (22) is true when the flag has a part that is green and a part that is white, and (23) is true when the dogs barked and the roosters crowed. Krifka's analysis of such sentences is based on Link's (1983; 1984) proposal for non-Boolean predicate conjunction. Link's motivation comes from relative clauses with more than one head noun, which he called "hydra constructions". Examples are in (24) and (25).

(24) The boy and the girl who met yesterday are friends of mine
(25) Students and workers who had gathered issued a resolution

In (24) and (25), a Boolean analysis according to which we distribute the relative clauses over the conjuncts does not work. As a solution, Link introduces a weaker definition of predicate conjunction, which is described in (26).

(26) A predicate P_1 and P_2 holds of x iff x can be divided into two parts x_1 and x_2 such that $P_1(x_1)$ and $P_2(x_2)$ hold

In order to account for the non-Boolean behavior of (24) then, Link proposes that the plurality x to which the predicate *boy and girl* applies is a sum of two entities - x_1 and x_2 - such that *boy* (x_1) and *girl* (x_2) hold, and that the sum x is the plurality that is taken to be in the extension of the predicates *met yesterday* and of *are friends of mine*.

Krifka generalizes Link's non-Boolean conjunction to other domains, and bases his analysis for conjoined adjectives and conjoined verbs on Link's definition of predicate conjunction for hydrae. Krifka (1990) proposes that any predicate conjunction P_1 and P_2 holds of an entity x if x can be partitioned into two parts x_1 and x_2 such that P_1 holds of x_1 and P_2 holds of x_2. For example (22) above, this means that *the flag* can be partitioned into two parts, where the predicate *green* holds of one of these parts and the predicate *white* holds of the other. The same reasoning holds for (23), where *the dogs and the roosters* is partitioned into the dogs (x_1) and the roosters (x_2) and *barked* holds of x_1 while *crowed* holds of x_2. In general, this means that on the basis of examples that show clear non-intersective behavior, Krifka assumes a weak, non-intersective meaning of *and* in general. His account, however, does not clearly specify how and when *and* receives an intersective interpretation.

Winter (2001a) acknowledges that while Krifka's non-intersective approach is a proper analysis for sentences like (22) and (23), it fails to capture the fact that other sentences only have an intersective interpretation. Winter provides examples of conjunction that he claims to show clear intersective behavior, as in (27), and contrasts these with minimally different examples that are not interpreted intersectively, as in (28).

(27) The ducks are swimming and quacking
(28) The ducks are swimming and flying

In order to account for the difference between pairs like (27) and (28), Winter (2001a) claims that on top of Krifka's descriptive proposal of non-intersective conjunction, we also need a principle that determines when which analysis is allowed. Winter (2001a) proposes that a maximality principle like Dalrymple et al.'s (1998) Strongest Meaning Hypothesis (SMH) is a suitable candidate for selecting either an intersective or a non-intersective analysis of conjunction. First, he assumes that the SMH is not construction-specific to plural sentences with reciprocals – as I presented it in Chapter 2, following Dalrymple et al. (1998). Instead, Winter (2001a) rephrases it into a general principle of plural predication, such that "any complex plural predicate with a meaning that is derived from one or more singular predicates using universal quantification is interpreted using the logically strongest truth conditions that are not contradicted by known properties of the singular predicate(s)" (Winter, 2001a). Note that unlike Dalrymple et al.'s SMH, Winter's extended SMH (eSMH) does not speak of 'context' in general, but focuses on a more manageable part of context, namely the lexical information that is tied to predicates. The contrast between minimal pairs like (27) and (28) is then captured in the following way. The eSMH selects the logically strongest possible candidate meaning for each sentence. When a strong interpretation (intersective conjunction) is consistent with properties of the predicates, then this is the attested interpretation of the sentence – an example is sentence (27). Only when such a strong interpretation is inconsistent with these properties, the interpretation is weakened. We see this in sentence (28): an intersective interpretation in which all ducks are in the intersection of the set of swimming individuals and the set of flying individuals contradicts what we know about swimming and flying. Thus, sentence (28) receives a split interpretation, which is the strongest interpretation that does not contradict this knowledge.

To summarize, previous works have either used the existence of clear non-intersective examples to argue for a non-Boolean analysis (Link, 1983; 1984; Krifka, 1990) or used the existence of clear intersective examples to argue for saving a Boolean analysis (Winter, 2001a). I claim that these views that only rely on simple contrasting examples do not sketch the full picture, hence do not account for the data reported in this chapter.

3.3 Conjunction of plural adjectival and verbal predicates: new data

This subsection summarizes new experimental data that are problematic for previous accounts. These data reveal effects that could not be shown reliably through informal collection of speaker judgments alone. Any previous theory of conjunction will need to be stated differently in order to account for these observations. After presenting the main data, I end the subsection with an overview of the remaining problems, indicating the need for the proposal that will be outlined in section 3.4.

Examples in favor of non-Boolean conjunction such as the ones in (22)-(25) have in common that, according to our world knowledge, the two conjoined predicates cannot apply to one individual simultaneously. For example, a flag cannot be green and white (without assuming different parts), dogs cannot bark and crow, and an individual cannot be both a boy and a girl. The incompatibility of the predicates forces a non-Boolean analysis of the conjunction, and hence sentences with such incompatible predicates are expected to be acceptable in a split situation. On the other hand, examples in favor of Boolean conjunction such as the one in (27) contain conjoined predicates that very naturally apply to one individual simultaneously. For example, it is very natural for a duck to be swimming and quacking, i.e. the predicates are compatible. Consequently, the sentence is expected to be acceptable in a joint situation. However, the data presented in this chapter reveal that if we look at more examples, strict compatibility vs. incompatibility of predicates are not the relevant notions. In fact, compatibility of predicates turns out to be a graded notion, and this graded compatibility allows us to make more fine-grained predictions about the interpretation of sentences. I will show that an objective measurement of compatibility correlates with the acceptability of a sentence in a split situation. In short, the data are as follows.

3.3.1 The new data

My experiments (described in detail in sections 3.5 and 3.6) include plural sentences with conjoined verbal predicates as in (29)-(30) and conjoined adjectival predicates as in (31)-(32).

(29) The men are sitting and reading
(30) The men are sitting and standing
(31) The animals are big and gray
(32) The animals are big and small

The predicate pairs in (29) and (31) are intuitively compatible: they can naturally apply to one individual simultaneously. By contrast, the predicate pairs in (30) and (32) are incompatible: according to world knowledge, they cannot apply to one individual simultaneously. Sentences with predicates that we consider incompatible (e.g. (30) and (32)) are highly acceptable in a split situation in which one of the predicates holds of a part of the plural individual and the other predicate holds of the rest (81% for adjectives and 84% for verbs). By contrast, sentences with predicates that we consider compatible (e.g. (29) and (31)) are less acceptable in a split situation (15% for adjectives and 54% for verbs). These results reveal that the logical interpretations of minimally different plural sentences with predicate conjunction differ, depending on different lexical information that is tied to the predicates: apparently, our world knowledge about verbs like *sitting* and *standing* as well as adjectives like *big* and *small* affects the way that we interpret the conjunctive sentences.

If we would take into account only the contrast in acceptability between sentences with compatible and incompatible adjective pairs in the split situation (15% vs. 81%), this could be considered experimental evidence in favor of Winter's (2001a) eSMH. However, when we also consider the data on sentences with verb pairs (54% vs. 84%), we see that it is clearly not the case that intuitively incompatible predicates lead to 100% acceptance of the split situation and intuitively compatible predicates lead to 0% acceptance – going against Winter's predictions. Thus, the knowledge affecting interpretation cannot merely comprise information about predicates being compatible or incompatible (0 or 1). Instead, using an independent measure of compatibility, I show that it contains more fine-grained conceptual information in terms of typicality preferences – just like with reciprocals in Chapter 2. This becomes even clearer when we also consider sentences with conjoined verbal predicates in which the predicates possibly, but not typically apply to one individual simultaneously. The compatibility test that I report shows that the compatibility value for such predicates is somewhere between 0 and 1. Examples are in (33) and (34).

(33) The men are sitting and cooking
(34) The men are waving and drawing

For example, it may be exceptional, but it is possible for a person to wave and draw simultaneously. My experiments show that sentences like (33) and (34), with verbs of a moderate compatibility value, are accepted in split situations by many speakers (78% of the time). Most importantly, to account for these data, I show that the degree to which a sentence with conjoined verbs is acceptable in a split situation correlates with the compatibility value of the two conjoined verb concepts.

3.3.2 Remaining problems

Now that I have summarized the main new data, I can objectively evaluate both Krifka and Winter's proposals. I claim that both solutions to non-intersective conjunction of predicates in plural sentences are not sufficient in accounting for the new data that is presented in this chapter.

If we adopt Krifka's (1990) non-intersective analysis without further assumptions, then all sentences with predicate conjunction (e.g. (29)-(34)) are expected to be 100% acceptable in a split situation. This prediction is clearly not borne out by my experimental data. Krifka's proposal alone does not account for the fact that subjects reject sentences in a split situation, nor that they reject them to different degrees depending on lexical information. If Krifka, however, would assume that the meaning of *and* is pragmatically strengthened to set intersection whenever possible (making it the reverse proposal to Winter (2001a)), it would still be difficult to account for the fact that sentences like (33) and (34) are accepted 78% of the time in a split situation. Since an individual man is able to sit and cook at the same time, or wave and draw at the same time, one would expect the interpretation of *and* in (33) and (34) to be pragmatically strengthened so that (33) entails that every man is both sitting and cooking, and (34) entails that every man is both drawing and waving. Following the same reasoning, the finding that sentences with supposed compatible verbal predicates like (29) are accepted in a split situation 54% of the time is a problem for Krifka's analysis. In conclusion, Krifka's proposal requires a mechanism of pragmatic strengthening. Without specifying the workings of such a mechanism, we cannot explain the data. This does not mean that Krifka's proposal is wrong, only that it is incomplete. In fact, the proposal that I put forward in section 3.4 could in principle be implemented into Krifka's original idea of a general weak meaning of conjunction.

Winter's extended SMH is more explicit with respect to when intersective interpretations are expected, and when in turn non-intersective interpretations are expected – making it possible to test its predictions. However, Winter's proposal also does not account for the acceptability of (29), (33) and (34) in split situations. If Winter is correct in assuming that non-intersective interpretations are only available when intersective interpretations are strictly ruled out by the predicates that are being conjoined, then all three of these sentences would have the same meaning: an intersective one in which every man is involved in both activities. The reasoning is that each man is physically able to be involved in both activities that are described in the sentences simultaneously, hence an intersective interpretation of these sentences is not ruled out, and a non-intersective interpretation should not be available. My data go against this prediction. As mentioned, examples like (33) and (34) are in fact very often accepted in a split situation – in which a non-intersective interpretation applies. Moreover, even examples with naturally compatible predicates like (29) are accepted in split situations to a relatively high degree. That means that the truth-conditions required by the eSMH, just like the original SMH for reciprocal sentences, are too strong. Moreover, the eSMH does not account for the significant difference in acceptability of sentences like (33) in a split situation compared to sentences like (29). The eSMH is unable to distinguish between them, i.e. between the combinations *sitting and reading* and *sitting and cooking*. I claim that this shortcoming has exactly the same nature as the SMH's shortcomings with reciprocals, where it could not distinguish between the concepts *know* and *pinch*, and consequently not between reciprocal sentences containing the concepts *know* and *pinch*. I will show that in order to understand the acceptability patterns for plural sentences with conjoined predicates, we also need a mechanism that has recourse to fine-grained conceptual information. I do not necessarily argue against previous proposals, but aim to show that they need to be stated differently, based on experimental work that tests the role of typicality effects. To that end, I put forward a reformulation of the Maximal Typicality Hypothesis.

3.4 The Maximal Typicality Hypothesis

In this section, I will argue that the Maximal Typicality Hypothesis – as was introduced in Chapter 2 in the context of reciprocal sentences – can also be applied to plural sentences with predicate conjunction. Using the MTH, I provide a solution to the problems for the eSMH by accounting for the acceptability patterns of triplets such as (29), (30) and (33) – repeated below as (35)-(37).

(35) The men are sitting and reading
(36) The men are sitting and standing
(37) The men are sitting and cooking

As expected based on results in Chapter 2, I will argue that these problems appear because there are certain aspects of meaning that the eSMH does not take into account. Based on the findings on reciprocal sentences in that chapter, I expect that we can overcome this shortcoming by taking typicality effects in categorization (see subsection 2.3.2) to be part of the contextual information that affects sentence interpretation.

3.4.1 The MTH: Connecting typicality with the interpretation of predicate conjunction

For plural sentences with predicate conjunction, taking typicality into account means that we take a more fine-grained notion of context into account than the eSMH. This is exactly what also appeared necessary for the analysis of reciprocal sentences. Here again, in addition to the definitional aspects that the eSMH considers (can a given situation be categorized as an instance of a verb concept *X*?), the MTH also has recourse to aspects of typicality (what preferences between situations does a verb concept *X* induce?). As we saw, the sharp distinction that Winter's (2001a) eSMH makes between possible and impossible situations, forces it to apply a Boolean analysis to both sentences (35) and (37). This is incorrect for (37). Under the MTH, the interpretation of the two sentences differs due to differences in typicality information between the concept combinations *sit* and *read* vs. *sit* and *cook*.

I apply the MTH to plural sentences with predicate conjunction in the following way. Just like for reciprocal sentences, the MTH uses information about typicality of different situations with respect to given predicate concepts. In sentences with predicate conjunction, the MTH uses typicality information on two conjoined concepts *P1* and *P2*. Based on that information on *P1* and *P2*, the MTH defines one situation as the *core situation* that is described by a sentence containing *P1* and *P2*. In the core situation, the agents in the sentence perform "enough" of the two actions described by the conjuncts to satisfy the conjunctive operator *and*.[20] For sentence (35), with the verbs *sit* and *read*, only a joint situation in which each man is sitting and each man is reading has enough actions. By contrast, for sentences (36) and (37), a split situation already has enough actions. I use the MTH to define exactly what is meant by "enough for the conjunctive operator". The MTH submits that the differences between sentences (35)-(37) follow from the typicality differences between the verb concept combinations. With the combinations *sit* and *stand* or *sit* and *cook*, adding more actions to a split situation would reach a situation where at least one individual is both sitting and standing/cooking. As we will see, this is atypical for the verb concepts, and hence the split situation is considered to already have "enough" actions for (36) and (37). The combination *sit* and *read*, however, allows the joint situation without any reduction of typicality. Thus, only the joint situation attains enough actions for (35). Similar to what we saw with reciprocal sentences in Chapter 2, the difference between these sentences is accounted for by the observation that the split situation is the maximal situation among the most typical situations for the verb concept combinations *sit* and *stand*, and *sit* and *cook*, but not for *sit* and *read*. Accordingly, the MTH selects the split situation as the core situation for (36) and (37), but not for (35). The process of selecting the core situation for a sentence with conjoined predicates on the basis of typicality information that is conveyed by the main concepts is defined below.

Maximal Typicality Hypothesis (MTH): For a plural sentence with two predicate concepts *P1* and *P2* in the scope of the conjunctive operator, situation S_C is the core situation for the sentence iff S_C is maximal among the situations that are most typical for *P1* and *P2*.

Once we determine which situations are most typical for predicate concepts that are conjoined within a plural sentence, the MTH selects the core situation for that sentence. For selecting the core situation among the most typical situations, I use the notion of 'maximal situation'. This notion relies on our ability to order situations according to containment relations – similar to what we saw for reciprocal sentences in Chapter 2. For example, a joint situation in which each of four individuals is both sitting and reading properly contains a situation in which only two of four individuals are sitting and two are reading. Consider the situations in Figure 1, in which activities are distributed over individuals in different ways.

[20] For simplicity, I assume here that the conjuncts describe 'actions'. This is obviously not true for adjectives such as *tall* and *thin* in *The men are tall and thin*. For such sentences, the core situation contains "enough" attributed properties described by the conjuncts. The reasoning regarding what is considered "enough" is similar to what we describe for 'actions', hence the contrast is not relevant for the point I make here.

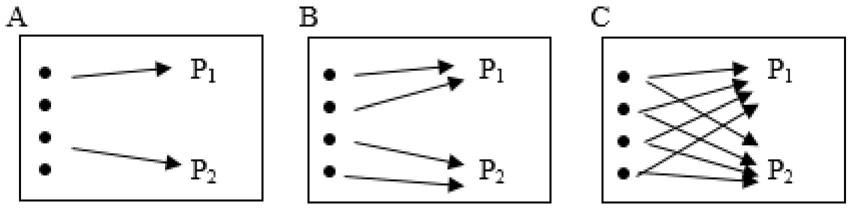

Figure 1: Three situations with four individuals

Figure 1 schematically represents three different situations. Each of the situations includes four individuals, represented by dots, plus two activities denoted by the predicates, represented by P_1 and P_2 (e.g. the verbs *sit* and *read*). The arrows represent the activities that the individuals are involved in. Figure 1A represents a situation in which only two of the four individuals are active: one individual is involved in activity 1 and another individual is involved in activity 2. Such a situation could be called an 'incomplete split situation'. It is 'split' in the sense that P_1 applies to one individual and P_2 applies to another individual, and it is 'incomplete' in the sense that not all four individuals are involved. Figure 1B represents a situation in which two of the individuals are involved in activity 1 and the other two individuals are involved in activity 2. This is the situation that I have referred to as the 'split situation'. Finally, figure 1C represents a situation in which each individual is involved in both activity 1 and activity 2. This is what I called the 'joint situation'. The joint situation properly contains the split situation, which properly contains the incomplete split situation. Accordingly, among these three situations, the joint situation is maximal, since it contains both the split situation and the incomplete split situation, and is not contained by any of them. Using this notion of containment, the MTH selects the core situation for each plural sentence with predicate conjunction. In sentence (35), where I assume that all three situations in Figure 1 are equally typical for the concepts *sit* and *read*, the joint situation is the maximal among them, hence it is selected by the MTH. By contrast, in sentences (36) and (37), I assume that the incomplete split situation and the split situation are most typical for the verb concepts, whereas the joint situation is less typical (or even impossible). Hence, the maximal situation among the most typical situations is now the split situation, and accordingly, this is the situation that is selected by the MTH as the core situation for (36) and (37).

Next, I use the MTH for explaining the acceptability pattern that I observe for sentences such as (35)-(37) in all situations, including non-core situations. The reasoning is the same as for reciprocal sentences. Firstly, sentences with conjunction of plural predicates are always expected to be highly acceptable in the core situation that the MTH specifies. Secondly, just like with reciprocal sentences, there are two possible types of situations that differ from the core situation:

a) *Situations that are properly contained in the core situation* (i.e. those that contain fewer actions): in those situations, sentences are expected to be less acceptable than in the core situation, with decreasing acceptability the fewer actions there are. For example, if the MTH selects the joint situation as the core situation for a sentence (e.g. sentence (35)), then that sentence is predicted to be less acceptable in the split situation than in the joint situation, and less acceptable in the incomplete split situation than in the split situation.

b) *Situations that properly contain the core situation* (i.e. those that contain more actions): in those situations, there are "more than enough" actions to satisfy the conjunctive operator. Whether such situations are relevant for actual use of sentences with conjoined predicates is determined by whether speakers categorize such situations as possible instances of the relevant verb concepts. For example, if the split situation is specified as the core situation for both sentences (36) and (37), then both sentences are predicted to be true and highly acceptable in this situation. However, with respect to the joint situation, judgments are affected by whether the joint situation is judged as a possible instance of the verb concept combinations. For *sit* and *cook*, the joint situation is a possible instance of the concepts. Accordingly, speakers are expected to have judgments on sentence (37) in the joint situation, and to judge the sentence as highly acceptable. By contrast, for *sit* and *stand*, the joint situation is unlikely to be judged as a possible instance of the concepts. Accordingly, sentence (36) will be much less acceptable in the joint situation, not because it is strictly speaking false, but because speakers are expected not to have judgments on (36) in the joint situation.

3.4.2 Approximating typicality

In line with the work that was described in Chapter 2, in order to test the MTH we need information regarding typicality of situations like the ones in Figure 1 for combinations of predicate concepts. The experimental work aimed at testing the MTH will focus on conjunction of verbal predicates in particular. For a verb concept like *sit*, I assume typicality effects much like for the concept *bird*: within the instances that are categorized as 'sitting' instances (or members of the *sit* category), I predict that some are consistently judged more typical than others. For example, a situation in which a person is sitting straight up in a chair is probably judged as a more typical instance of *sit* than a situation in which a person is leaning so far back that they are almost lying down on the floor. For the particular situations that I am interested in (Figure 1), however, directly measuring typicality is quite a complicated, or at least strange task. Consider having to judge the typicality of the situations depicted below in Figure 2 for the concept *sit*.

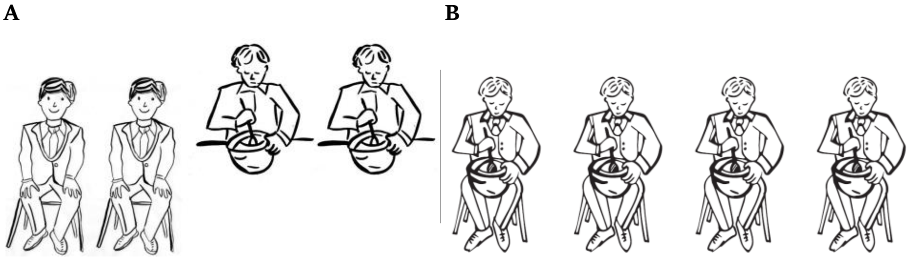

Figure 2: Two possible situations for the concept *sit* and the concept *cook*: split situation (A) and joint situation (B)

In the situation in Figure 2A, two men are sitting while the other two men are cooking (split situation), while in the situation in Figure 2B, every man is both sitting and cooking (joint situation). In its account of plural sentences with predicate conjunction, the MTH relies on the assumption that such situations may differ with respect to how typical they are for the concept *sit* and the concept *cook*. Since it is difficult to measure such typicality effects immediately for concepts like *sit* and *cook*, and even impossible for concepts like *sit* and *stand* (since they cannot be depicted in a joint situation), I rely on a simplified, indirect measurement for the typicality of situations as in Figure 2. Intuitively, we might think that situation A is a more typical instance of *sit* than B, and that it is also a more typical instance of *cook* than B. This is because in situation B, the fact that each man is involved in both activities causes the typicality of the situation as an instance of each concept in isolation to decline. A situation in which one is sitting while cooking is probably not the most typical instance of *sit*, and definitely not a typical instance of *cook*. My simplified measurement indirectly measures the typicality of joint situations (e.g. B in Figure 2) in a textual test. That measurement tests how typical it is for two verbs to apply to one individual simultaneously, i.e. their 'compatibility'. This allows me to directly compare different pairs of verbs. Ideally, if one were to be interested in the full typicality structure of verb concepts like *sit*, one could construct a standard task (similar to the tasks used for noun concepts like *bird*), namely rating all possible *sit* situations with respect to how typical they are for the concept. In the current work however, the measurements are restricted since they are led by a direct research question, based on an observation: I am specifically interested in the different interpretation patterns of sentences like (35)-(37), thus looking for a measurement that allows me to directly compare 'sitting and cooking' versus 'sitting and reading' versus 'sitting and standing'. One should keep in mind, however, that when I speak of 'compatibility' of predicate concepts P_1 and P_2, I aim to indirectly measure the typicality of an instance of P_1 and the typicality of an instance of P_2. The reason I did not test this visually, i.e. presenting subjects with a situation like B in Figure 2 in which both verbs apply and measure its typicality as an instance of verb concept P_1 and of verb concept P_2 (similar to measuring the typicality of an *ostrich* as an instance of the concept *bird*), was because I was also interested in strictly incompatible pairs of verbs (e.g. *sit* and *stand*) – which cannot be depicted within one situation. So instead, I conducted a more indirect, simple textual compatibility test, in which I assess the typicality of a joint situation for verb concept P_1 and verb concept P_2 by measuring the verbs' compatibility. If two verbs are incompatible, I infer that split situations (Figure 1B) are more typical for the verb concepts than joint situations (Figure 1C). I use this information for evaluating the MTH.

3.4.3 Predictions of the MTH: sentence interpretation

Now that I have discussed how I measure typicality effects, let me illustrate the precise empirical predictions of the MTH for the interpretation of sentences (35)-(37). Given the verbs *sit* and *read* in (35), my experimental data reveal that these verbs are highly compatible, i.e. subjects consider it natural if both verbs apply to one individual simultaneously. That means that the fact that both verbs apply simultaneously (i.e. a person is reading while sitting, or sitting while reading) does not affect the typicality of the situation for each verb concept in isolation. Based on this measurement, I extrapolate that the situations in Figure 1 (incomplete split, split and joint) are equally typical for the verb combination *sit* and *read*. Using this extrapolation, the MTH predicts that the core situation that (35) describes is the joint situation: the maximal situation among the three situations in Figure 1. Thus, I predict that (35) is fully acceptable in the joint situation, which agrees with the predictions of the eSMH. Furthermore, I predict (35) to be less acceptable in the split situation, and even less so in the incomplete split situation, because these situations are properly contained in its core situation.

For *sit* and *stand* in sentence (36), my data reveal that they are highly incompatible, i.e. subjects consider it strange if both verbs apply to one individual simultaneously. The fact that both verbs apply simultaneously seriously affects the typicality of the situation for each verb concept in isolation. In fact, such a situation is probably physically impossible.[21] Assuming a threshold model (Hampton, 2007), we could say that it falls below the threshold for category membership of *sit* or *stand* (when the typicality of a situation increases for the verb concept *sit*, it decreases for the verb concept *stand*, and vice versa). I extrapolate from this that in the most typical situations for the verb concept combination *sit* and *stand*, there is no individual that is involved in both activities. The MTH predicts that the core situation described by (36) is the maximal one among these situations: the split situation, in which each individual

[21] Or, if you could imagine some strange situation of sitting and standing simultaneously, then it would at least be a highly atypical instance of *sit*. This does not affect the nature of the argument.

is involved in exactly one activity.[22] Adding more actions to such a situation would result in a situation that is not among the most typical ones. Hence, (36) is predicted to be fully acceptable in the split situation. This is in line with the predictions of the eSMH. Sentence (36) is expected to be less acceptable in the incomplete split situation than in the split situation, because the incomplete split situation is properly contained in the core situation. Finally, I predict (36) to be unacceptable in the joint situation because, despite the fact that it is technically true, the joint situation is unlikely to be judged as a possible instance of the concepts *sit* or *stand*.

Finally, the example that is most problematic for previous proposals is sentence (37). The core situation that is selected by the MTH does not coincide with the strongest possible interpretation that is selected by the eSMH. A joint situation, where every man is sitting and every man is cooking, is physically possible for sentence (37). Therefore, according to the eSMH, this situation is expected to be the only situation in which (37) is true. By contrast, the MTH predicts (37) to be true in the split situation. As I will show, the verbs *sit* and *cook* are judged compatible only to a certain degree, somewhere between *sit* and *read*, and *sit* and *stand*. A situation in which a person is simultaneously sitting and cooking, however odd, can be categorized as an instance of the concept *sit* (i.e. it is above the category membership threshold) and the concept *cook*. Interestingly however, the fact that both verbs apply at the same time, causes the typicality of the situation as an instance of each verb concept in isolation to decline for most participants – hence leading to a moderate compatibility value. A situation in which one is sitting while cooking is probably not the most typical instance of *sit*, and definitely not a typical instance of *cook*. From the typicality measurement for such verbs, I extrapolate that for most participants the most typical situation for the verb combination *sit* and *cook* is the split situation. For that majority of the participants, (37) is acceptable in the split situation. I also predict (37) to be true in the joint situation, which properly contains the core situation. Unlike (36), (37) is expected to be acceptable in the joint situation, because the concepts *sit* and *cook* allow the joint situation as a possible instance of the concepts.

In order to test these predictions, I measured typicality of the joint situation for predicate concepts (via compatibility) and the interpretation of sentences containing those concepts.

---

[22] Technically, under this reasoning, there are other possible core situations for a sentence like (36). The core situation could also be a situation in which one predicate applies to all individuals and the other predicate applies to none of the individuals (e.g. four men are sitting and no men are standing) or a situation in which one predicate applies to three individuals and the other predicate applies to only one of the individuals (e.g. three men are sitting and one man is standing). However, I consider a split situation in which both predicates apply to a plurality of two individuals as an optimal split situation. I believe that this is not due to typicality, but for independent reasons (e.g. Gricean reasoning (Grice, 1975)).

3.4.4 Overview of experimental work

The remainder of this chapter reports two experiments with two separate goals: Experiment 2 and the main experiment, Experiment 3. Before testing the predictions of the MTH, Experiment 2 experimentally tests the non-trivial predictions of the simpler hypothesis: Winter's (2001a) eSMH. The results of Experiment 2 combined with previous results of Experiment 1 in Chapter 2 – which revealed the effects of typicality on reciprocal interpretation according to the MTH - led me to conduct the more sophisticated Experiment 3. The goal of Experiment 3 was to test whether verbal predicate conjunctions with atypical pairs of predicates reveal a challenge for the eSMH, and whether the MTH makes the right predictions for these cases. Note that the experiments reported in this chapter do not compare acceptability of sentences in different situations (split/joint) as in Chapter 2 (*S3*/*S6*), but only compare the acceptability of sentences in a split situation. This is due to presentational issues: we cannot depict the joint situation for sentences with incompatible predicates, such as *big and small* or *sit and stand*.[23]

Experiment 2 tested the predictions of the eSMH on sentences with adjectival predicate conjunction in a truth-value judgment task. The eSMH predicts that sentences with predicate conjunction receive an intersective interpretation, unless such an interpretation is strictly excluded by properties of the conjoined predicates. Testing those predictions only required contrasting the interpretation of sentences with compatible predicates to sentences with incompatible predicates. Unlike with verbs, with adjectives introspection is a good tool to construct clearly compatible pairs (e.g. *The animals are big and gray*) and strictly incompatible pairs (e.g. *The animals are big and small*). For that reason, Experiment 2 is a quick and simple first test that did not require any pretests or sophisticated measures. I compared the acceptability of sentences with naturally compatible adjectives vs. strictly incompatible adjectives in the split situation. If the eSMH is correct, sentences should only be accepted in the split situation if the adjectives strictly exclude the joint situation.

The next step was the more elaborate Experiment 3, which included combinations of predicates with varying degrees of compatibility. Experiment 3 systematically tested the predictions of the MTH for plural sentences with conjoined verbal predicates. The set-up of the experiment was similar to Experiment 1, which tested the MTH for plural sentences with reciprocal expressions. In order to test the relationship between typicality and interpretations, I made use of two separate parts. Part 1 tested interpretation, specifically the acceptability of plural sentences with conjoined verbal

[23] Type 3 verbs (e.g. *bite*) in Chapter 2 in principle had the same presentational issue in situation *S6*, i.e. the *S6* situation could not be depicted for reciprocal sentences like *John, Bill and George are biting each other*. In Chapter 2 we circumvented this problem by adopting a schematic method. For the purposes of the current chapter, I also explored the possibility of a schematic method, but such a method proved to be more difficult than the one I eventually used and report here.

predicates in a split situation. Part 2 tested typicality effects for verb concept combinations through a measurement of ‘compatibility’. That way, I measured how natural a joint situation was judged for a given pair of verbs. If the MTH is correct, sentences with atypical verb combinations should be highly acceptable in a split situation in Part 1, because that situation is the core situation based on the measurement of typicality in Part 2.

3.5 Experiment 2: Testing the extended SMH with adjective conjunction

Experiment 2 focused on the interpretation of Dutch plural sentences with conjoined adjectival predicates in order to test the predictions of the eSMH. It measured the acceptability of such sentences in a split situation.

Participants

Twenty-nine students from Utrecht University (27 female, age M = 21) participated for monetary compensation. All participants were native speakers of Dutch without dyslexia.

Materials

Twenty-six different Dutch sentences with conjoined adjectives were tested on their acceptability in a split situation, in pen-and-paper questionnaires. I used two different questionnaires, each containing 13 unique test items plus 13 filler items that were the same across the two questionnaires. Each test item contained a plural sentence with adjective conjunction of the form *The X are P_1 and P_2* (where X is a plural noun and P_1 and P_2 are adjectives), plus an image depicting four entities (people, animals, objects or shapes) in a split situation. In that situation, predicate P_1 always applied to entities A and B, and predicate P_2 applied to entities C and D. An example of a test item image is given in Figure 3.

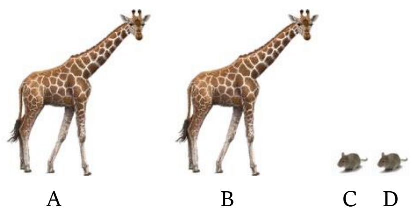

Figure 3: Stimuli from Experiment 2: example of a test item image for the sentences *The animals are big and small* and *The animals are big and gray*.

The independent variable that was manipulated is the compatibility of adjectives P_1 and P_2 in the test sentence. In each questionnaire, half[24] of the test items contained sentences with adjectives that are lexically incompatible (e.g. *The animals are big and small*), and the other half contained sentences with predicates that are lexically compatible (e.g. *The animals are big and gray*). The incompatible pairs were created by combining adjectives that cannot apply to a given individual entity simultaneously, for example size adjectives that are antonyms (e.g. *big and small*) or shape adjectives (e.g. *triangular and rectangular*). The compatible pairs were created by combining adjectives that can naturally apply to a given individual entity simultaneously, for example a size adjective and a color (e.g. *big and gray*) or a shape adjective and a color (e.g. *triangular and black*). The same 13 test item images were used for both compatible and incompatible sentences. To ensure that participants never saw the same image twice, one of these sentences occurred in version 1 and the other occurred in version 2. This allowed me to compare acceptability of minimally different sentences given the same images. The items of each version are represented by light and dark cells in table 1 (for the original Dutch material see table B1 in Appendix B).

Table 1: Overview of adjective pairs, translated from Dutch
(light and dark cells represent the division of items over the two versions of Experiment 2).

Entities	Compatible	Incompatible
Animals	Big and gray	Big and small
Objects	Long and black	Long and short
Objects	Striped and red	Striped and plain
Objects	High and red	High and low
People	Redhaired and angry	Redhaired and blondhaired
People	Short and blackhaired	Short and tall
People	Blue-eyed and blackhaired	Blue-eyed and brown-eyed
People	Thin and glasses-wearing	Thin and fat
People	Bald and angry	Bald and with-hair
People	Longhaired and glasses-wearing	Longhaired and shorthaired
Shapes	Solid and red	Solid and dotted
Shapes	Triangular and black	Triangular and rectangular
Shapes	Round and green	Round and square

Filler items also contained images with four entities, but a different type of accompanying sentence without predicate conjunction. The accompanying sentences in the filler items were either sentences with quantifiers (e.g. *Most boys are P*) or

[24] Since the number of test items was uneven in each questionnaire, the division was seven compatible vs. six incompatible or seven incompatible vs. six compatible.

sentences mentioning only two or three of the entities in the picture (e.g. *A, B and C are P*). Half of the filler items were expected to be judged true, and half of them were expected to be judged false. Both versions of the questionnaire contained the same filler items.

Procedure

Each participant completed one of two questionnaires on paper, in a sound-proof booth. They were instructed to indicate as soon as possible whether they judged the sentence to be true or false given the image, by circling "true" or "false" with a pen.

Coding and analysis

Responses were coded '1' when participants judged a sentence to be true for a given image, and '0' when they judged a sentence to be false. I computed the proportion of true-responses for each of the two types of sentences (compatible and incompatible) for each participant. I then performed a repeated measures ANOVA with Compatibility as the within-subjects factor (with 2 levels: compatible and incompatible) and Version as between-subjects factor (with 2 levels: version 1 and version 2).

Results

Results of Experiment 2 revealed differences between the two different types of items. Sentences with compatible adjectives were accepted in a split situation only 15% of the time, while sentences with incompatible adjectives were accepted 81% of the time (see Table B1 in Appendix B and Figure 4 for detailed results).

Unsurprisingly then, the repeated measures ANOVA revealed a significant main effect of Compatibility (F (1,27) = 91.74, p < .001). As expected, the results show that sentences with incompatible predicates are accepted more often in the split situation than sentences with compatible predicates.

Furthermore, there is no effect of Version (F (1,27) = 2.67, p = .11) nor an interaction effect of Version*Compatibility (F (1,27) = .17, p = .68), indicating that the two versions of the questionnaire do not differ from each other, and thus measure the same thing.

Discussion

Experiment 2 aimed to test the predictions of Winter's (2001a) eSMH, which phrases Dalrymple et al.'s original SMH as a more general meaning principle that applies to other domains besides reciprocals. Specifically, the experiment tested whether the interpretations of plural sentences with conjoined adjectival predicates are sensitive to the compatibility of the adjectives in the sentence. The eSMH predicts conjunction in such sentences to always be interpreted intersectively, except when lexical properties of the predicates strictly exclude such an interpretation. In that case, the interpretation is weaker and consequently the sentence is expected to be accepted in a split situation.

In order to test this hypothesis, I measured the acceptability of sentences with either lexically compatible or lexically incompatible conjoined adjectives in a split situation: a situation with four individuals in which one of the adjectives applies to two individuals and the other adjective applies to the two other individuals.

I found a clear difference in acceptability between sentences with different pairs of adjectives: those with incompatible pairs (e.g. *big and small*) showed much higher acceptability in the split situation than sentences with compatible pairs (e.g. *big and gray*). This result indicates that compatibility of conjoined predicates indeed affects the meaning of the conjunction, and provides experimental support for Winter's proposal of a more general SMH.

Interestingly though, although there is a large difference in acceptability between sentences with compatible and incompatible pairs in the expected direction, the difference in not 100% vs. 0% acceptance. Moreover, even though all sentences with incompatible pairs in Experiment 2 showed significantly higher acceptability in a split situation than sentences with compatible pairs, we still see some variation between items (Figure 4).

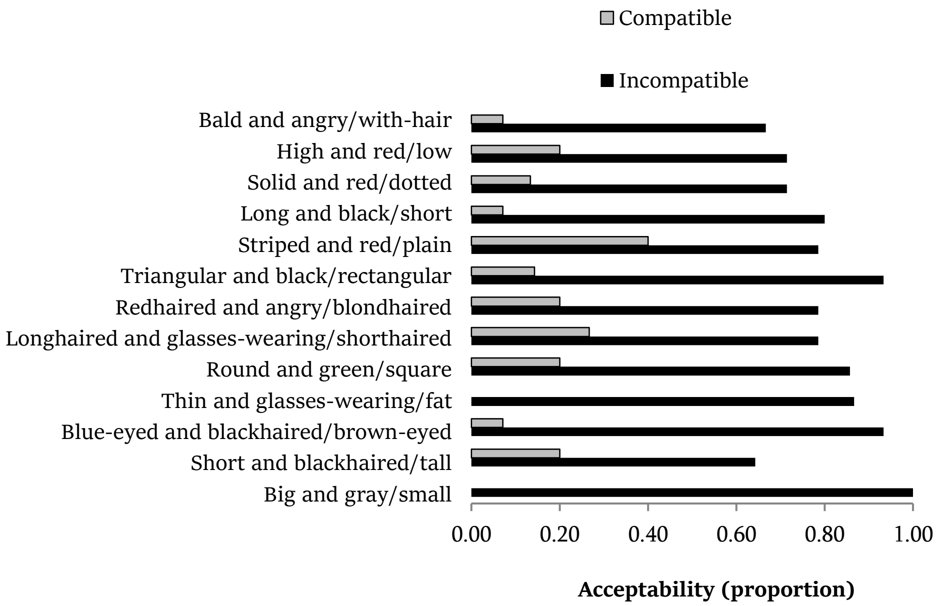

Figure 4: Mean acceptability per item per compatibility type (translated from Dutch)

The stimuli for Experiment 2 were selected based on an a priori assumption that they were either incompatible adjectives or compatible adjectives. However, the difference among the items suggests that compatibility might not be an absolute, clear-cut matter. Those differences in acceptability of different pairs in split situations can be explained if compatibility is not just a matter of 'yes' or 'no'. In order to get more insight into the relationship between different combinations of predicates in terms of compatibility on the one hand, and interpretation of sentences on the other hand, we need to test compatibility of predicate concepts independently.

I expect that degrees of compatibility exist, though the eSMH does not make any predictions based on such degrees. However, the Maximal Typicality Hypothesis that we used to explain the data on reciprocal sentences in Chapter 2, uses exactly such fine-grained information about concepts. I claim that this is the key to explaining the problematic examples that we discussed earlier, e.g. (37), and also the key to explaining the variation across items in Experiment 2. If we extend the MTH to the domain of predicate conjunction, we can make clear and testable predictions just like we did for reciprocal sentences. Unlike in the case of reciprocals where the relevant typicality parameter was patient cardinality, here the theory needs to have recourse to the typicality of two predicates applying simultaneously to one individual – their compatibility value.

I predict that independent typicality measurements would explain why the items in Experiment 2 slightly differ, namely because their typicality preferences (i.e. their compatibility value) also differ. The adjectives that we initially called 'compatible' but in fact show quite high acceptability in the split situation (i.e. *striped and red*) might not be 100% compatible after all. They are expected to be slightly less compatible than those that show zero acceptability (e.g. *big and gray*). Experiment 3 systematically tested this prediction.

3.6 Experiment 3: Testing the MTH with verb conjunction

3.6.1 Introduction

Experiment 3 tested the predictions of the MTH for plural sentences with conjoined predicates. Unlike Experiment 2 which tested sentences with conjoined adjectives, Experiment 3 tests sentences with conjoined verbs. The reason for moving from adjectives to verbs is due to experimental considerations. Verbs, unlike adjectives, easily allow the construction of stimuli with possible but atypical combinations. It is easy to come up with possible but atypical verb pairs that apply to humans by introspection, for example *sit* and *cook* or *walk* and *read*, but much more difficult to come up with possible but atypical adjective combinations. To illustrate, it is difficult to come up with a possible yet atypical size and color combination for an animal, or a possible yet atypical hair color and eye color combination for a human.[25] For more discussion on this point see subsection 3.7.1.

In order to test the predictions of the MTH, Experiment 3 made use of two separate parts. Part 1 measured the acceptability of sentences with pairs of conjoined verbs in a split situation. Part 2 measured typicality effects for verb concept combinations by means of measuring their compatibility value given one individual that performs two actions simultaneously. From that I extrapolate how typical a joint situation is for a given pair of verbs. I expected sentences such as *The men are sitting and cooking* to be highly acceptable in a split situation in Part 1 because that is the core situation: it is maximal among the situations that are most typical for the conjoined concepts - as independently measured in Part 2. Hence, the MTH is evaluated on the basis of a correlation analysis between the two parts.

Materials for Experiment 3 were constructed based on pretests that were conducted in order to include a wide range of compatibility values in the actual experiments. In order to test the predictions of the MTH, I aimed to include compatible and incompatible pairs of verbs, but also pairs of verbs that can possibly apply to one individual simultaneously, but typically do not.

3.6.2 Pretests: Constructing materials

Pretest 1

The aim of the first pretest was to gather as many Dutch verb combinations as possible, especially atypical ones. Eight participants were provided with 16 sets of two pairs of predicates, *P_1 and P_2* and *P_1 and P_3*: one very natural pair, and one pair that is physically impossible to apply simultaneously, e.g. *sitting and reading* (*P_1 and P_2*) and *sitting and standing* (*P_1 and P_3*). I then asked participants to provide as many verbs that they could come up with that combine with *P_1* (i.e. *sitting* in this case) that led to a possible but atypical, uncommon or strange pair. The pairs that participants constructed, combined with more natural pairs that I came up with, led to a list of 91 verb combinations in total.

Pretest 2

In the second pretest, 29 different participants rated all of these 91 pairs for compatibility, in a paper-and-pencil task. For each pair, participants were asked to rate how odd[26] they would consider it if both verbs applied to one person at the same time.

[25] I did consider continuing the work in this direction, using atypical but possible adjective conjunctions such as non-stereotypical combinations for humans and animals. However, after consulting a few speakers, I concluded that with verbs it is much easier to collect such atypical pairs.

[26] Phrasing the question negatively by asking subjects ‘how odd’ they would rate a situation was done because a) directly asking for ‘how compatible’ they would judge two predicates seemed like a too technical and too direct task, and b) asking for ‘how typical’

Oddness was rated on a 6-point scale, where 1 meant ‘not odd at all’ and 6 meant ‘physically impossible’. I mentioned explicitly that 5 thus meant ‘very odd, but physically possible’, in order to distinguish large atypicality from impossibility.

Results of this pretest showed great variability in ratings between verb pairs, with a high level of agreement between the participants (Cronbach’s alpha was .88 for the 91 items). The selection of verb pairs that were to be used in the main part of Experiment 3 proceeded as follows. I defined sets of verb pairs on the basis of the different P_1 verbs, e.g. a set consisted of *sitting and reading*, *sitting and standing, sitting and knitting, sitting and cooking,* etc. Then I selected the 12 sets that showed the greatest range of ratings. Finally, I selected three verb pairs from within each of these 12 sets: the verb pair that was rated lowest on the oddness scale (compatible pairs like *sitting and reading*), the verb pair that was rated highest (incompatible ones like *sitting and standing*), and a verb pair that was rated in between, at a mean of 4 points[27] (atypical pairs like *sitting and cooking*). The 36 verb pairs that constituted the final material, translated from Dutch, are given in Table 2 (the original Dutch material can be found in Table B2 of Appendix B). Creating the three groups (with labels ‘compatible’, ‘incompatible’ and ‘atypical’) was done purely to ensure variability while constructing the materials – similar to our labelling of type 1, 2 and 3 verbs in Experiment 1 on reciprocals. I will refer to these three groups when discussing set-up and subresults of Part 1 and Part 2 of Experiment 3. Note however that the distinction between the groups is not meaningful in the final correlation analysis of all data points.

---

they would judge a situation turned out to be ambiguous in Dutch. Some subjects interpreted the word *typical* to mean ‘atypical’, whereas asking for oddness is unambiguous.

[27] Additional inclusion criteria included that a) each verb should be expressed by one word only, b) ratings for verb pairs should have small variation (whenever there was more than one candidate for selection, the one with the lowest standard deviation for the ratings was selected). Finally, if after considering these criteria there were still two candidate pairs for the atypical group, we decided that c) atypical verb pairs should have no 6 point ratings (since that meant that at least one participant judged it to be physically impossible for the two verbs to apply simultaneously). This was only a very small criterion, applying only to one case.

Table 2: Overview of verb pairs, translated from Dutch
(light and dark cells represent the division of items over the two versions of Experiment 3).

Compatible	Incompatible	Atypical
Sitting and reading	Sitting and standing	Sitting and cooking
Waving and smiling	Waving and clapping	Waving and drawing
Walking and singing	Walking and swimming	Walking and writing
Crawling and screaming	Crawling and jumping	Crawling and reading
Standing and reading	Standing and squatting	Standing and falling asleep
Reading and smiling	Reading and sleeping	Reading and drawing
Lying down and stretching	Lying down and running	Lying down and drinking
Drawing and yawning	Drawing and typing	Drawing and walking
Swimming and smiling	Swimming and crawling	Swimming and reading
Texting and frowning	Texting and knitting	Texting and waving
Knitting and singing	Knitting and clapping	Knitting and walking
Sleeping and drooling	Sleeping and telephoning	Sleeping and standing

3.6.3 Experiment 3 (Part 1): Interpretation of plural sentences with verb conjunction

Part 1 of Experiment 3 measured the acceptability of 36 Dutch plural sentences with two conjoined verbs in a split situation. Each sentence was of the form *The X are P_1 and P_2* (where *X* is a plural noun and P_1 and P_2 are verbal predicates).

Participants
A total of 33 students from Utrecht University (28 female, age M = 21) participated for monetary compensation. All participants were native speakers of Dutch without dyslexia. Prior to the experiment all participants signed an informed consent form.

Materials
Thirty-six different Dutch sentences with conjoined verbal predicates were tested on their acceptability in a split situation. I used two versions of a truth-value judgment task in Dutch, each containing 18 unique test items plus 18 filler items that were the same across the two versions. Each test item contained a plural sentence with verb conjunction (*The X are P_1 and P_2*)[28] and a drawn image depicting four individuals in a

[28] All the sentences in the experiment were in the simple present tense, which can be used to describe ongoing events as well as states in Dutch. Whereas in English one would use the progressive tense for all sentences in Part 1 of Experiment 3, the distribution of the progressive tense in Dutch is different, such that it could not be used for all sentences in the experiment alike. For example, the progressive tense can quite naturally be used in Dutch

split situation. In that situation, predicate P_1 applied only to persons A and B, and predicate P_2 applied only to persons C and D. Half of the images depicted male individuals, and the other half depicted female individuals. Examples of test images are given in Figure 5.

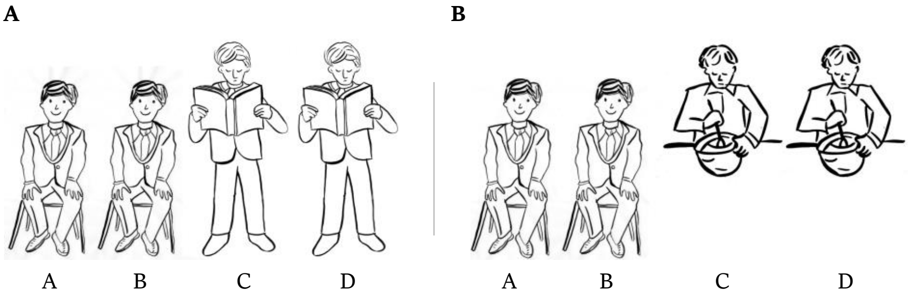

Figure 5: Stimuli from Experiment 3 (Part 1): examples of test item images for the sentences *The men are sitting and reading* and *The men are sitting and standing* (A) and *The men are sitting and cooking* (B) (translated from Dutch).

The independent variable that was manipulated is the compatibility of verbs P_1 and P_2 in the test sentence. Unlike in Experiment 2, this variable now had three levels: compatible, incompatible and atypical. In each version of the experiment, one third of the test items contained sentences with verb pairs that were considered compatible P_1 and P_2 in pretest 2 (e.g. *The men are sitting and reading*), one third contained sentences with verb pairs that were considered incompatible P_1 and P_2 (e.g. *The men are sitting and standing*) and one third contained sentences with pairs that were considered atypical P_1 and P_2 (e.g. *The men are sitting and cooking*). The same images were used for sentences with compatible and incompatible pairs with identical P_1 (e.g. Figure 5A for *The men are sitting and standing* and *The men are sitting and reading*).[29] To ensure

for action verbs like *cook* (*Ze zijn aan het koken*, 'They are cooking') but not for sentences with stative verbs like *stand* (*?Ze zijn aan het staan*, 'They are standing').

[29] Note that alternatively to using the same images for compatible and incompatible items, I could have used different images. In the current work, the decision was made to go for the first option, but I acknowledge that either option has its own advantages and disadvantages. On the one hand, using the same images for compatible and incompatible items might seem like making the task more complex. Having the same images means that in the incompatible condition, two individuals are performing an extra action that is not mentioned in the sentence (e.g. the two persons on the right in Figure 5A are reading while this is not mentioned in *The men are sitting and standing*), which might cause some confusion. On the other hand, having the same images for both sentences makes the comparison 'cleaner', i.e. without any additional contextual differences between the conditions. Furthermore, it makes it more comparable to the method used in Experiment 2. The alternative option of having separate images for the conditions might also be worthwhile, but I leave it for follow-up research.

that participants never saw the same image twice (such as the one in Figure 5A), one of these sentences occurred in version 1 and the other occurred in version 2. The atypical items (e.g. Figure 5B for *The men are sitting and cooking*) were divided over the two versions, resulting in two experiments with six sentences with compatible pairs, six sentences with incompatible pairs and six sentences with atypical pairs each, accompanied by 18 unique images. The items of each version are represented by light and dark cells in table 2.

Filler items contained similar images with four people, but a different type of accompanying sentence. The accompanying sentences in the filler items were either sentences with quantifiers (e.g. *Some boys are P*) or sentences mentioning specific individuals in the picture (e.g. *Boys A, B and C are P*). Half of the filler items were expected to be judged true, and half of them were expected to be judged false. Both versions of the experiment contained the same filler items.

The order of items was pseudo-randomized using Mix software (Van Casteren & Davis, 2006), with the following restrictions: items containing the same verb were at least six items apart; there were at most two test items immediately following each other, and at most two filler items immediately following each other; similar test items (in terms of compatible/incompatible/atypical) or similar filler items (in terms of quantifier/specific individuals) never immediately followed each other. Finally, I constructed two orders of each version, with the second one having reversed order of items.

Procedure

Each participant completed one version of the experiment. The task was presented in a sound-proof booth on a PC using Open Sesame software (Mathôt, Schreij & Theeuwes, 2012). Prior to entering the sound-proof booth, each participant received verbal instructions explaining the experimental set-up. Further, more detailed instructions were given on the PC monitor.

After being instructed, each participant completed three practice trials. Subsequently, they were given the opportunity to ask for clarifications, if necessary. No verb used in the practice session appeared in the actual experiment. The experiment itself consisted of the 36 items described above. Image and sentence were presented in the center of a white screen. Participants were instructed to indicate as soon as possible whether they judged the sentence to be true or false given the situation in the image by pressing the left or right button with their dominant hand.

Coding and analysis

Responses were coded '1' when participants judged a sentence to be true for a given image, and '0' when they judged a sentence to be false. The proportion of true-responses were computed for each of the three types of sentences (compatible, incompatible and atypical) for each participant. I then performed a repeated measures

ANOVA across participants with Compatibility as the within-subjects factor (with 3 levels: compatible, atypical, and incompatible).[30] Post-hoc Bonferroni corrected multiple comparisons were performed in order to analyze differences between different Compatibility levels in detail. An ANOVA across items, with Compatibility as the between-item variable (also with 3 levels), gave similar results to the participant analysis. Therefore I only report the first analysis.

Results

The results of Part 1 of Experiment 3 are summarized in Table 3. More detailed results on acceptability per item are in Table B2 of Appendix B. Overall, the truth percentages of the different sentences in the experiment ranged from 24% to 100%. I predicted lowest acceptability for the sentences with compatible pairs and highest acceptability for the sentences with incompatible pairs.

Table 3: Mean acceptability in Part 1 of Experiment 3.

Compatibility type	% "true" judgments (st. dev.)
Compatible	54 (31)
Atypical	78 (31)
Incompatible	84 (32)

A repeated measures ANOVA revealed that there was a main effect of Compatibility (F (1.36, 43.49) = 37.41, p < .001). Pairwise comparisons reveal that all three levels differ significantly from each other in acceptability: the acceptability of sentences with compatible verbs differs from the acceptability of sentences with incompatible verbs (p < .001); the acceptability of sentences with compatible verbs differs from the acceptability of sentences with atypical verbs (p < .001); and the acceptability of sentences with incompatible verbs differs from the acceptability of sentences with atypical verbs (p < .05). Note again, however, that the main conclusion from this experiment is not that there are significant differences between groups but the fact that there is variation in the data, i.e. that sentences with different predicate combinations result in different percentages of acceptability.

[30] A repeated measures ANOVA with Version as between-subjects factor was also performed, but showed no effect of Version (F (3, 29) = .47, p = .71) nor an interaction effect of Version*Compatibility (F (6,58) = .82, p = .58). I thus collapsed the versions for the analysis.

3.6.4 Experiment 3 (Part 2): Compatibility of verb pairs

Part 2 of Experiment 3 measured the compatibility value for the 36 Dutch verb concept pairs that were used in sentences of Part 1. The aim was to measure the typicality of one particular situation as an instance of the concepts P_1 and P_2, namely one in which both verbs *P_1 and P_2* apply simultaneously (the joint situation). As discussed in subsection 3.4.2, I conducted an indirect textual test in which the typicality of the joint situation for P_1 and P_2 is assessed by measuring the predicates' compatibility. This test is a replication of the pretest, but carried out by different participants and now containing fewer items, in a more controlled experiment with pseudo-randomized orders.

Participants
The same 33 students from Utrecht University from Part 1 participated in this experiment. Each subject completed Part 1 first, before proceeding with Part 2. Also, in between experiments they took part in a third, unrelated experiment.

Materials
I used a questionnaire containing 36 Dutch statements about one person involved in two actions simultaneously. Half of the statements were about males and half of them were about females (matching the gender of persons in the pictures of Part 1). Each statement contained a singular object (a male or a female) and two conjoined verbs *P_1 and P_2* (e.g. *The man is sitting and reading*). The 36 pairs of verbs were the same as the ones used in sentences of Part 1, thus one third of the pairs were considered compatible in the second pretest (e.g. *sitting and reading*), one third were considered incompatible (e.g. *sitting and standing*), and one third were considered atypical (e.g. *sitting and cooking*).

The order of items was pseudo-randomized using Mix software (Van Casteren & Davis, 2006), with the restrictions that items containing the same P_1 never immediately followed each other, and at most two items of the same type (in terms of compatible/incompatible/atypical) immediately followed each other.

Finally, four different orders of the questionnaire were constructed: two versions that started with the statements about males (with the second one having reversed order within males and females statements), and two versions that started with the statements about females (with the second one having reversed order within males and females statements).

Procedure
Each participant received one of the questionnaires on paper, in a sound-proof booth. They were instructed to rate how odd they would consider it if both verbs applied to the given person at the same time (see footnote 26). Oddness was rated on a 6-point

scale, where 1 meant ‘not odd at all’ and 6 meant ‘physically impossible’. We mentioned explicitly that 5 thus meant ‘very odd, but physically possible’, in order to distinguish large atypicality from impossibility.

Coding and analysis

Responses were coded ‘1’ through ‘6’ corresponding to the participant’s oddness judgment. This way the incompatibility rating for each verb pair was computed. I performed a repeated measures ANOVA across participants with Compatibility as the within-subjects factor (with 3 levels: compatible, atypical, and incompatible). Post-hoc Bonferroni corrected multiple comparisons were performed in order to analyze differences between different Compatibility levels in detail. An ANOVA across items, with Compatibility as the between-item variable (also with 3 levels), gave similar results to the participant analysis. Therefore I only report the first analysis.

Results

Results of Part 2 of Experiment 3 are summarized in Table 4. More detailed results on incompatibility rating per item are in Table B2 of Appendix B. Overall, mean ratings per verb pair ranged from 1.03 to 5.94, and there was a very high correlation between these ratings and the ratings for these items in the pretest ($r = .98$, $p < .001$).

Table 4: Mean incompatibility ratings in Part 2 of Experiment 3.

Compatibility type	Mean incompatiblity (st. dv.)
Compatible	1.35 (.29)
Atypical	3.82 (.58)
Incompatible	5.66 (.38)

As expected from the correlation with the pretest, the repeated measures ANOVA revealed that there was a main effect of Compatibility (F (1.95, 62.45) = 1187.02, $p < .001$). Pairwise comparisons reveal that all three levels differ significantly from each other: the incompatibility of supposed compatible pairs differs from the incompatiblity of supposed incompatible pairs ($p < .001$); similarly for the incompatiblity of compatible vs. atypical pairs ($p < .001$); and the incompatiblity of incompatible vs. atypical pairs ($p < .001$). This means that the three groups that we selected based on the pretest were confirmed in Part 2 of Experiment 3 (with different participants and a subset of the stimuli).

3.6.5 Correlation between results of Part 1 and Part 2

The MTH predicts that the acceptability of split situations (as measured in Part 1) correlates with the compatibility rating for two verb concepts - as an indirect measurement of the joint situation for those concepts (as measured in Part 2). That way, the MTH aims to explain the interpretation pattern for plural sentences with conjoined atypical verb pairs. The prediction was borne out: I found a significant correlation between the results on Part 1 and Part 2 (r = .66, n = 36, p < .001, see Figure 6).

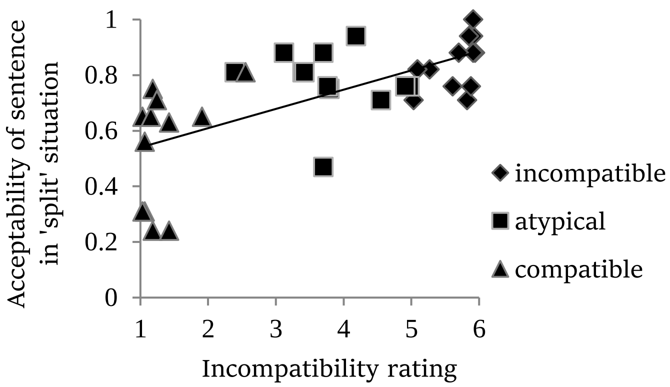

Figure 6: Relation between mean acceptability (in proportions of sentences in a split situation in Part 1 of Experiment 3 and mean incompatibility rating of verb pairs in Part 2 of Experiment 3.

Discussion

Experiment 3 tested the predictions of the MTH in the domain of plural sentences with conjoined verbs. The experimental set-up was similar to Experiment 1, which measured the predictions of the MTH for plural sentences with reciprocal expressions. The MTH specifies the core situation for a plural sentence as the maximal one among those situations that are most typical for the predicate concept(s) in the sentence. Based on that core situation, it makes detailed predictions about the acceptability of sentences in different situations. In order to test the hypothesis, Part 1 measured the acceptability of sentences with conjoined verbs in one particular type of situation – the split situation - and Part 2 measured the compatibility of two verb concepts (as an indirect test of the typicality of the joint situation for those concepts). To test the MTH, a correlation analysis on the two parts was performed. I predicted that there is a correlation between the acceptability of sentences in a split situation and the incompatibility of verb concepts in that sentence.

Part 1 revealed a continuum of acceptability values of sentences in a split situation, ranging from 24% to 100% acceptable. Sentences with incompatible verbs were accepted more often than sentences with compatible verbs, with mean acceptability values of 84% and 54% respectively. Sentences with atypical pairs of verbs were accepted 78% of the time in split situations. These results are not explained by Winter's (2001a) account. Since compatible pairs as well as atypical pairs of verbs do not strictly exclude a joint situation, the eSMH predicts sentences with such pairs to be downright unacceptable in a split situation. These findings are substantial support for the MTH over the eSMH, especially once we also consider the results of Part 2.

In Part 2, I found that the typicality of the joint situation differs for different verb concept pairs. This typicality was measured indirectly by measuring 'compatibility' of verb pairs: how typical it is for both verb pairs to apply to one individual simultaneously. Firstly, the verb pairs that were called 'compatible' indeed received a high compatibility value in Part 2 (i.e. a low incompatibility value). From that measurement, I extrapolated that joint situations are typical for such verb pairs. The MTH then predicts that the core situation for sentences containing compatible verb pairs is the joint situation, since this is the maximal situation. That means that sentences with compatible pairs are expected to be highly acceptable in the joint situation, and less acceptable in the split situation because the split situation is properly contained in the core situation, i.e. contains fewer distributed actions to the agents. I thus correctly predicted the acceptability of sentences with compatible verbs in the split situations to be lowest among the three groups in Part 1 of the experiment. Next, the verb pairs that were called 'incompatible' received a low compatibility value in Part 2 (i.e. a high incompatibility value). I extrapolated that the most typical situations for such verb pairs are those in which no individual is involved in two activities simultaneously. The MTH then predicts that the core situation for sentences containing incompatible verb pairs is the split situation, since this is the maximal situation among those that respect the verbs' typicality preference. That means that such sentences are expected to be highly acceptable in the split situation, which is in line with the findings in Part 1. Finally, the verb pairs that were called 'atypical' received a compatibility rating somewhere in between the compatible and incompatible verb pairs. From that, I extrapolated that for most participants, the most typical situations with such verb pairs are those in which no individual is involved in two activities simultaneously, similar to the incompatible verb pairs. The MTH thus predicts that for most participants the core situation for sentences with atypical verb pairs is the split situation.

Most importantly, independently of which group a verb pair belongs to, the MTH predicts that the lower the degree to which participants consider two verbs compatible, the more they will consider the joint situation as atypical, and hence the more they will accept a sentence in the split situation. This predicted connection between acceptability and typicality was supported by a significant correlation between the results of Part 1 and Part 2.[31]

---

[31] Since I have data from the same participants in Part 1 and Part 2, J. Hampton (personal communication) pointed out that it would be useful to supplement the group level data with an individual-based analysis. There are at least two possible ways to do this. Firstly, one could compare for each individual participant a) the mean incompatibility rating (from Part

3.7 General discussion points

The results from Experiment 2 and 3 raise a few general discussion points. Firstly, there is a difference between adjectives in Experiment 2 and verbs in Experiment 3 regarding the acceptability of sentences with compatible predicates in the split situation. Secondly, in light of reciprocal sentences being considerably less acceptable in situation *S2* than in *S3*, another discussion point is what happens with incomplete split situations for sentences with predicate conjunction. In particular, whether this could shed light on the nature of the low acceptability of reciprocal sentences in *S2*. Thirdly, there is the question of whether the MTH also applies to plural sentences with conjoined nominal predicates that were discussed in 3.2.1. And finally, I discuss the possibility that sentences in my experiments are sometimes accepted because they contain sentential conjunction in their deep structure, allowing a reference shift of the plural subject.

3.7.1 Compatible adjectives vs. compatible verbs

There is a clear difference in the acceptability pattern between plural sentences with conjoined adjectives (in Experiment 2) versus conjoined verbs (in Experiment 3). Specifically, sentences with conjoined adjectives that were grouped under 'compatible' are accepted in split situations only 15% of the time, while sentences with conjoined verbs that were grouped under 'compatible' are accepted 54% of the time. Sentences with adjectives and verbs that I called 'incompatible' do converge nicely with respect to their acceptability in a split situation (81% and 84%, respectively). This raises the question what explains the differences between compatible adjectives and compatible verbs. First of all, there was a difference in stimulus construction. While adjective pairs in Experiment 2 were constructed based on introspection, verb pairs in Experiment 3

---

2) for those verb pairs for which they deemed the sentence in Part 1 true, with b) the mean incompatibility rating (from Part 2) for those verb pairs for which they deemed the sentence in Part 1 false. According to Hampton's threshold model (2007; see section 2.3.1), the mean rating should be lower for those verb pairs where the sentence in Part 1 was deemed false compared to the rating for those verb pairs where the sentence in Part 1 was deemed true. In other words, low typicality leads to higher acceptability of a sentence in the split situation than high typicality. A second way to check the threshold model in an individual-based analysis would be to compare for each participant the incompatibility rating for any given verb pair to the incompatibility rating of any other verb pair (from Part 2). According to the threshold model, if the incompatibility of pair 1 is lower than of pair 2, then it can never be the case that a sentence containing pair 1 was deemed true in a split situation while a sentence containing pair 2 was deemed false. Both of these additional analyses would be interesting, and I leave them for future work.

were constructed based on pretests. Thus, there exists the possibility that the notion of compatibility was different between the judges for both experiments. More likely, however, the attested difference has to do with the difference between the nature of adjective concepts and verb concepts. Simply put, I speculate that it is easier to have fully compatible or incompatible adjectives than it is to have fully compatible or incompatible verbs. Adjectives like *big and small* are clearly incompatible due to being contradictory and *big and gray* are clearly compatible (for the tested sentence *The animals are big and gray* we even have elephants as an obvious natural example). By contrast, verbs allow more subtleties, as can most obviously be seen through the existence of many so-called atypical verb combinations. The reason for that might be that most verb concepts have more dimensions in categorization than most adjectives. Sassoon (to appear) claims something similar for the distinction between different types of adjectives: “classification under *long* is merely a matter of length, but classification under *healthy* is a matter of degree in a variety of dimensions, such as blood-pressure, cholesterol and blood-sugar level” (Sassoon, to appear). I speculate that something along those lines can also be said about the distinction between the adjectives in Experiment 2 and the verbs in Experiment 3. Intuitively, adjectives such as *tall* and *thin* are one-dimensional. If two adjectives are on the same ‘scale’, e.g. *tall* and *short*, then they must be incompatible. If they are on a different ‘scale’, e.g. *tall* and *thin*, then they are compatible. This makes it very easy to judge whether two of such adjectives are compatible or incompatible. Verb concepts, however, are likely to be much harder to judge. In order to decide whether two verbs are compatible, we probably need to mentally simulate events. As a result, verbs leave more room for different interpretations than adjectives. A verb can be compatible with another verb in one respect (or on one ‘dimension’), but much less compatible in another respect. For example, *lying down and stretching* is compatible if one were to assume that *lying down* means being part of an event in which one is awake and has the physical room to stretch (for example in a gym). By contrast, the verbs are much less compatible in case *lying down* means being part of an event in which one is tired and inactive, lying in a bed under the covers, or even sleeping. Such differences could cause verb pairs (as in Experiment 3) to be less clearly compatible than adjective pairs (as in Experiment 2).

3.7.2 Incomplete split situations and the nature of *S2*

Based on the core situation that the MTH specifies for a given sentence, acceptability patterns are predicted for that sentence in different situations. Among other things, I predict that sentences are less acceptable in situations that are properly contained in the core situation compared to the core situation itself, with decreasing acceptability the fewer relations or distributed actions there are. In Chapter 2, I determined that the core situation for sentences with type 1 verbs (e.g. *know*) is *S6*. Hence, they were expected to be less acceptable in *S3* than in *S6*, and even less so in *S2*. This prediction was borne out, but there was a peculiar aspect to this finding: while sentences were acceptable in *S3* roughly 50% of the time, they were downright unacceptable in *S2*. In Chapter 2, I speculated that this might be the case because the threshold for category membership for the particular verbs that we tested is between *S2* and *S3*. For different verbs, such as *stacked on top of*, the threshold might be lower, since a sentence like *The planks are stacked on top of each other* seems quite natural in *S2*. In Experiment 3, the parallel to type 1 verbs would be compatible verb pairs, in the sense that the core situation for those sentences is also the strongest possible situation: the joint situation. Sentences with conjoined compatible verbs are hence also expected to be less acceptable in weaker situations: the split situation and the incomplete split situation (see section 3.4.1 for graphs of these situations). Indeed, results of Experiment 3 reveal that sentences with compatible verbs are accepted in the split situation roughly 50% of the time – similar to reciprocal sentences with type 1 verbs in *S3*. The question one could raise is what happens with sentences with compatible verbs in the weakest, incomplete split situation. I currently do not have data on such situations, but one could easily imagine that for the tested verbs there is a similar threshold between split and incomplete split situations, just like between *S3* and *S2*. That means that the sentences we tested would also be downright unacceptable in an incomplete split situation. If there are verb pairs for which the threshold is below the incomplete split situation, this could reveal more about the nature of the findings on reciprocals. It is very likely that the empirical domain is limited, however, since it is hard to imagine such verb pairs.

3.7.3 The MTH and plural sentences with nominal conjunction

The experimental work in the current chapter focused on conjunction of adjectival predicates and conjunction of verbal predicates. A question that reasonably arises is whether the MTH also applies to plural sentences containing conjunction of nominal predicates. Based on indications from the literature on nominals, I assume that while the MTH is consistent with intuitions about such sentences, it does not surface there due to distinct, independent strategies that apply to nominal conjunction. Let us consider sentences (18) and (21) that I presented in section 3.2.1, repeated below as (38)-(39).

(38) Ten men and women got married today in San Pietro
(39) These linguists and philosophers all came to the party

As I discussed in 3.2.1, previous works on the topic have consistently argued that non-intersective, split interpretations are available for plural sentences with conjoined incompatible nominals as in (38) as well as with compatible nominals as in (39). While sentence (38) allows only such a split reading (due to the incompatibility of the predicates *men* and *women*), sentence (39) is ambiguous between a split reading and a joint reading. Thus, independently of the compatibility of the predicates, a split reading is assumed to be available for such sentences. Despite the fact that the predicates *linguists* and *philosophers* are perfectly compatible, i.e. *these linguists and philosophers* can refer to one group of individuals who are both linguists and philosophers, the split reading is still expected to be naturally available. My experiments do not provide data on nominal conjunctions, but it is to be expected that (39) would never be rejected in a split situation where the subject refers to a group of individuals, some of whom are linguists and some are philosophers. This is in contrast to the sentences with adjectival conjunction or verbal conjunction that I studied in Experiment 2 and Experiment 3, respectively. The interpretation of conjoined adjectives and conjoined verbs is dependent on the compatibility of the conjoined predicates. High compatibility leads to rejections of sentences in split situations: sentences with compatible adjectives are generally rejected in a split situation and sentences with compatible verbs are rejected roughly 50% of the time. I thus speculate that nominal conjunction involves a different phenomenon. For such sentences, the literature points out many examples of split interpretations with both compatible and incompatible predicates. If these are correct, that means that any proposal that generates a weak, split interpretation could be applied to plural sentences with conjoined nominals (e.g. Krifka, 1990; Heycock & Zamparelli, 2005). Consequently, a strategy that always generates a weak meaning disguises the effects of the MTH, i.e. we cannot see correlations between compatibility and interpretation.

In sum, I believe that there are different mechanisms at work in sentences with nominal conjunction. These mechanisms do not require a principle like the MTH, but they are consistent with it.

3.7.4 Reference shift of the plural subject?

The particular interpretation that was in the focus of Experiment 2 and Experiment 3 was the split interpretation in which one predicate applies to two of the individuals in the picture and the other predicate applies to the two remaining individuals. I have claimed that this interpretation is sometimes available for predicate conjunction sentences, namely to the degree that a situation in which the conjoined predicates apply simultaneously is atypical. One might argue instead that the acceptability of these sentences given a split interpretation has nothing to do with typicality. As an alternative, one might reason that we accept a sentence like (36), repeated below as (40) because its deep structure is the sentential conjunction in (41), which contains two definite plurals that hence allow the possibility of referring to two different groups of men. In other words, the reasoning would be that (40) is accepted in a split situation because we are able to very quickly shift the reference of the plural noun *the men* from one set of men to another set of men. My experiments would then in fact deal with reference resolution instead of with matters of typicality.

(40) The men are sitting and standing
(41) The men$_i$ are sitting and the men$_j$ are standing

If indeed the predicate conjunction in (40) is actually interpreted as sentential conjunction (as is made explicit in (41)), then I would expect to see no differences between different test sentences. If reference shift would explain why sentence (40) is accepted in a split situation, then this strategy should be available across the board for all types of predicate conjunction that were tested – whether they are typical, atypical or incompatible. This is clearly not the case, and the question remains what explains the range of acceptability values.

Tieu et al. (under revision) show that sentential conjunctions of incompatible predicates (e.g. *The bears are big and the bears are small*) are accepted in a split situation – suggesting that either reference shift of the plural subject is at work, or perhaps a principle like the MTH works at the sentential level. However, an additional pilot study[32] that I conducted - including both sentential conjunction of compatible predicates and incompatible predicates - has revealed that acceptability in the split situation is high for both types of sentences. Subjects accepted sentence (41) (with incompatible predicates) given a split situation, but they equally accepted sentences like (42) and (43) given a split situation – even though we have seen that sentences containing such compatible predicates behave differently when presented as mere predicate conjunctions (i.e. they are generally judged false 50% of the time). A single subject did not accept sentence (41) in a split situation, but they also refused to accept sentences like (42) and (43) in a split situation.

(42) The men are sitting and the men are reading
(43) The men are waving and the men are smiling

What this suggests is that reference shift is independent from the obtained results in Experiment 3, which support the MTH. Sentences with sentential conjunction do not show the same correlation with typicality as sentences with predicate conjunction do. I believe reference shift of the plural subject in sentences like (41)-(43) is purely motivated by trying to make a sentence true. This explains why most participants always accepted such sentences in a split situation (i.e. used reference shift to make the sentence true). The one subject that did not use reference shift, was consistent in

[32] The study was conducted with 9 participants who were students at Utrecht University (6 female, age M = 23), and checked the acceptability of 12 plural sentences with sentential conjunction in a split situation. Each sentence was of the form *The x are* P_1 *and the x are* P_2 (where x is a plural noun (used twice) and P_1 and P_2 are verbal predicates). Half of the P_1 and P_2 pairs were compatible predicates while the other half were incompatible predicates (based on pretests from Experiment 3). One participant accepted none of the sentences, the remaining eight participants accepted all or all but one.

not using it across different types of predicate conjunction (compatible and incompatible).

Another argument against the reference shift explanation is the finding (based on a small pilot[33]) that sentences with proper name conjunctions instead of definite plurals are accepted in a split situation significantly more often when the conjoined predicates are incompatible (as in (44)) compared to when they are compatible (as in (45)).

(44) John, Bill, Sue and Jane are sitting and standing
(45) John, Bill, Sue and Jane are sitting and reading

For such sentences, reference shift of the subject *John, Bill, Sue and Jane* is obviously not possible, and still a sensitivity to the predicate concepts in the sentence is observed, which is along the same lines as the results presented in the current chapter.

Summarizing, I conclude that it is unlikely that the presented results from Experiment 2 and Experiment 3 are due to reference shift of the plural subject. Shifting the reference of the subjects from one referent to another does not explain the systematic variability in acceptability of the split situation, nor that a similar pattern arises for sentences with proper name conjunctions as in (44) and (45).

3.8 Conclusion

This chapter first described some well-known problems for an intersection analysis of conjunction, which have led some researchers to propose a collective theory of conjunction while others have attempted to save an intersective account. By experimentally researching plural sentences with conjoined predicates, I have shown that neither of the proposals can do without fine-grained lexical information on content words. While Winter's (2001a) eSMH takes into account properties of the conjoined predicates, it does not make the correct predictions for sentences with atypical pairs of verbs: those that possibly but atypically apply to one individual simultaneously. I proposed that the Maximal Typicality Hypothesis that was introduced for reciprocal sentences in Chapter 2, is a suitable meaning principle for plural sentences with conjunction as well. With a set of experiments, I have shown that the MTH makes the correct predictions, by selecting the core situation for sentences based on typicality effects for the conjoined verbs – revealing a systematic advantage over the eSMH. More

---

[33] This study was conducted with 22 participants who were students at Utrecht University (18 female, age $M = 19$). It checked the acceptability of 8 plural sentences with predicate conjunction in a split situation. Each sentence was of the form *A, B, C and D are* P_1 *and* P_2 (where *A, B, C* and *D* are names and P_1 and P_2 are verbal predicates). Half of the P_1 and P_2 pairs were compatible predicates while the other half were incompatible predicates (based on introspection). Sentences with compatible pairs were accepted in a split situation 10% of the time, sentences with incompatible predicates were accepted in a split situation 40% of the time.

generally, the results imply that truth conditions of plural conjunctive sentences should be stated relative to a speaker's conceptual knowledge.

Chapter 4
Plural predicate conjunction as a window into conceptual combination in the brain

This chapter is a slightly expanded version of Poortman and Pylkkänen (2016).

Experiment 2 in Chapter 3 investigated the interpretation of plural sentences with two conjoined adjectives as in (1) and (2). Results of that experiment revealed that the logical interpretation of such sentences is dependent on the lexical information that is tied to the adjectives.

(1) The animals are gray and big
(2) The animals are small and big

As discussed in Chapter 3, Winter (2001a) predicts that sentences containing compatible adjectives such as (1) receive an intersective interpretation. That means that (1) is true iff the animals are gray and the animals are big. By contrast, sentences containing strictly incompatible predicates such as (2) receive a non-intersective, split interpretation under Winter's hypothesis. That means that (2) is true iff a subset of the animals is small and the remaining animals are big. Experiment 2 tested these predictions by measuring the acceptability of minimally different sentences such as (1) and (2) in a split situation. Results revealed that sentences with incompatible predicates were accepted in a split situation much more often than sentences with compatible predicates – in line with the predictions. In the current chapter, we[34] use this finding in a neurolinguistic study on a specific brain region that has been implicated in the composition of complex concepts: the Left Anterior Temporal Lobe

[34] The research in Chapter 4 was carried out together with Liina Pylkkänen. It was supported by the National Science Foundation Grant BCS-1221723 (L.P.) and grant G1001 from the NYUAD Institute, New York University Abu Dhabi (L.P.).

(LATL). Using the observed differences between sentences with compatible vs. strictly incompatible, contradictory adjectives allowed us to further specify the processes that occur in this region. Results of Experiment 2 indicated certain cases of adjective conjunction as optimal test cases: those that are always interpreted intersectively and those that are always interpreted non-intersectively. In a sentence with naturally compatible predicates which is always interpreted intersectively such as (1), both properties of being big and being gray are attributed to all individual members of the set of animals. By contrast, in a similar sentence with incompatible predicates which is never interpreted intersectively such as (2), the properties of being big and being small are attributed to the set, but not to each individual in that set. Thus, in terms of concept composition, we could say that in a sentence which is interpreted intersectively, two 'features' are combined to form a single coherent entity representation, resulting in more specific conceptual representations (e.g. big gray animals) than in a sentence which is not interpreted intersectively (e.g. big animals, small animals). Based on results from Experiment 2, we were able to construct pairs of syntactically identical sentences that differ only with respect to the attribution of their denoted properties to the members in the set denoted by the subject. Employing such minimal pairs allowed us to test specifically whether the LATL region is sensitive to the number of properties that is integrated across different members of a set.

MEG

The method that we use is magnetoencephalography (MEG). MEG allows non-invasive measurement of ongoing brain activity, with both high spatial resolution and high temporal resolution. This is in contrast to more common measures like fMRI and EEG, which provide precise spatial information or timing information, respectively, but not both. MEG measures brain activity by recording the magnetic fields that occur whenever electric current flows within neurons. When individual neurons in the brain are active, this results in a flow of electrically charged ions through the cell. When multiple neurons are active at the same time in a specific area, this generates electromagnetic fields that can be recorded outside of the head.

4.1 Introduction

A fundamental aspect of human language is the productive composition of complex expressions from simpler ones. For example, a proficient speaker of English can rapidly and effortlessly combine individual words like *man* into phrases such as *tall man*, and multiple phrases into sentences such as *the tall man saw a dog*. A central goal for the neuroscience of language is uncovering the brain bases of linguistic composition, thus unpacking the specific computations underlying the composition of complex meaning. Previous research has identified the left anterior temporal lobe (LATL) as a key contributor to sentence processing in general (e.g. Mazoyer et al., 1993; Stowe et al., 1998; Friederici et al., 2000; Brennan & Pylkkänen, 2012), and to the basic combinatory operations of language more specifically, as for instance exemplified by the composition of small adjective-noun phrases like *red boat* (e.g. Bemis & Pylkkänen, 2011; 2013a, b). More recent MEG work aimed at uncovering the particular aspects of composition that the LATL supports, suggests that it is specifically involved in the composition of complex concepts, rather than syntactic or semantic composition in a more general sense (Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015; Del Prato & Pylkkänen, 2014), in line with literature implicating the LATL for aspects of semantic memory (e.g. Baron & Osherson, 2011). The MEG work has shown that the LATL combinatory effect is sensitive to the conceptual specificity of the composing items (Westerlund & Pylkkänen, 2014) and that it is absent for cases of numeral quantification (e.g. *two cups*), where the first word of the phrase does not add a feature to the noun but rather enumerates the number of tokens within a set (Del Prato & Pylkkänen, 2014). It thus appears that the computation that is relevant for the LATL is some process in which a conceptual feature X is added to a representation Y.

To further specify the computational contribution of the LATL to combinatory processing, Experiment 4 tested whether LATL effects as observed in MEG require a situation in which features such as 'big' and 'gray' combine to form a single coherent entity representation (as in (1), but not in (2)) or whether the relevant computation simply requires the attribution of such features to a set but not necessarily to the same members of the set (not distinguishing between (1) and (2)). Under the former hypothesis, the LATL would be sensitive to the number of features added to the representation of a single entity whereas under the latter account, LATL activity would reflect the total number of features integrated across different members of a set. To test this, we employed conjunctions of two adjectives as in Experiment 2, whose lexical semantics were varied such that they either naturally allowed or strictly disallowed the attribution of their denoted properties to the same members of a set, i.e. the properties were either compatible (*gray and big*) or strictly incompatible (*small and big*).

The chapter is structured as follows. The remainder of this section describes previous works that aimed to uncover the precise workings of the LATL in language processing, and introduces the new work in this chapter. Section 4.2 reports on Experiment 4: a new MEG study on adjective conjunction and the LATL. Finally, 4.3 and 4.4 are discussion and conclusion, respectively.

4.1.1 Prior evidence for composition and conceptual processing in the LATL

Several literatures have implicated a combinatory role for the LATL. Within hemodynamic research, many studies have compared the processing of full syntactically well-formed sentences to the processing of unstructured word lists

(Mazoyer et al., 1993; Stowe et al., 1998; Friederici et al., 2000; Vandenberghe et al., 2002; Humphries et al., 2005, 2006; Jobard et al., 2007; Rogalsky & Hickock, 2009; Pallier et al., 2011). These studies systematically showed an increased LATL response for the sentences over the unstructured word lists, indicating that the LATL is involved in sentence processing. Crucially, however, the processing of a full sentence is a complicated endeavor, involving a multitude of sub-processes, such as syntactic parsing of the sentence structure, semantic composition of the sentence meaning, pragmatic inferencing, and so forth. Thus, several different hypotheses regarding the function of the LATL remain consistent with the sentence vs. list findings.

Later work using MEG has aimed to narrow down this hypothesis space using very minimal experimental designs (Bemis & Pylkkänen, 2011; 2013a, b). These studies moved away from the sentence vs. list paradigm, and instead compared the processing of simple adjective-noun combinations (e.g. *red boat*) to one-word stimuli (e.g. *boat*) or two-word lists (e.g. *cup*, *boat*). Results of these studies consistently showed more activity in the LATL for the adjective-noun condition than for their non-combinatory controls. These findings suggest that the LATL effects observed in sentence processing studies reflect basic combinatorial operations as opposed to other components of sentence processing. A remaining question within this line of work, however, is whether the LATL is responsible for syntactic or semantic composition, since the processing of adjective-noun combinations clearly involves both.

In answer to this, subsequent research zooming in even further on the role of the LATL has shown that its activity is unlikely to reflect any general form of either syntactic or semantic composition. Instead, it appears more specifically involved in the composition of some type of complex concepts (Westerlund & Pylkkänen, 2014; Del Prato & Pylkkänen, 2014). This hypothesis is in line with literature on the 'hub and spoke' model of conceptual knowledge (e.g. Patterson et al., 2007; Lambon Ralph et al., 2010). The 'hub and spoke' model characterizes the LATL as a semantic, amodal hub connecting a widely distributed neural network of areas that each contain modality-specific features of concepts (including sensory, motor and verbal areas). The hub serves as an additional layer that integrates the different types of conceptual information in order to build an amodal representation of a concept. Consistent with this, damage to ATL regions correlates with a disorder called Semantic Dementia, which affects conceptual knowledge. Categorization performance of patients with Semantic Dementia is modulated by concept specificity: specific concepts, with more distinctive features, are more affected than less specific ones (Warrington, 1975; Hodges et al., 1995; Done & Gale, 1997; Rogers et al., 2004, 2006). This pattern can be predicted by the 'hub and spoke' model if it is assumed that the hub is responsible for categorization based on featural overlap. To investigate the relationship between combinatory effects and conceptual specificity effects in the LATL, Westerlund and Pylkkänen (2014) compared low and high specificity nouns (e.g. *boat* vs. *canoe*) in both combinatorial and non-combinatorial contexts (e.g. *blue boat/blue canoe* vs. *xhsl boat/xhsl canoe*). They found that low specificity nouns exhibited a larger amplitude increase as a result of adjectival modification than high specificity nouns, i.e. the LATL showed a larger combinatory effect for lower specificity items. Thus, it is possible that the specificity boost provided by adjectival modification was larger when the adjective modified a less specific noun, a generalization also supported by a subsequent finding where more specific modifiers (such as *vegetable* in *vegetable soup*) were found to elicit larger combinatory effects on the noun than less specific ones (Zhang & Pylkkänen, 2015). These findings are unpredicted by functional hypotheses associating the LATL to syntactic or semantic composition in general terms, as these computations should not be sensitive to conceptual specificity, and instead conform to an account in which the LATL more specifically reflects the construction of complex concepts. The goal of Experiment 4 was to further elucidate what variety of conceptual combination is relevant for the LATL.

4.1.2 Current study: Intersective vs. collective conjunction as a window to LATL function

In the studies summarized above, composition or conceptual combination was always achieved by modifying nouns with adjectives, as in *red boat*. Another way to add features to a concept is via conjunction: *this box is red and large*. The extant literature on the LATL would strongly predict the LATL to be sensitive to this type of conceptual combination. However, conjunctions do not necessarily add features to a single individual, instead, they may also distribute the introduced features to different members of a set. In Chapter 3, I showed that adjective conjunction can be interpreted intersectively and non-intersectively. Consider again sentences (1) and (2), repeated below.

(1) The animals are gray and big
(2) The animals are small and big

Sentence (2) can be used to truthfully describe a situation with a set of small animals (e.g. mice) and a set of large animals (e.g. elephants): a non-intersective, split situation. Crucially, the two conjoined predicates *small and big* are incompatible and hence do not add features to the same individuals within the set of animals. In contrast, sentence (1) is generally interpreted intersectively, such that it describes a situation with a set of animals who are gray and big (e.g. a set of elephants). The conjoined predicates *gray and big* add the features of grayness and bigness to all individual members in the set of animals. Note that in Chapter 3 I showed that mere ‘incompatibility’ and ‘compatibility’ are not the correct notions when accounting for the full interpretation pattern of plural sentences with predicate conjunction. While acknowledging that, Experiment 4 in this chapter only includes those cases of conjunction that are always interpreted intersectively (e.g. *gray and big*) and those that are always interpreted non-intersectively (e.g. *small and big*) - as measured in a truth-value judgment task. To keep the contrast between these interpretations clear throughout this chapter, we shall call the latter a **collective** interpretation, as it attributes both properties to a set collectively but not to the individuals (e.g. Krifka, 1990; Heycock & Zamparelli, 2005; Winter, 2001a). The former interpretation will simply be called **intersective** (e.g. Keenan & Faltz, 1985; Partee & Rooth, 1983).

If the notion of "conceptual combination" that is relevant for the LATL involves the combination of features to form a single coherent entity representation (which of course could be a plurality as in (1)), then only intersective interpretations should engage the LATL. This is what a 'hub and spoke' model most naturally predicts: since intersective conjunction adds features to the same individuals, it serves to further specify those individuals' representation, i.e. *the animals are gray and big* results in a more specific animal-representation than the single property predication in *the animals are big*. In contrast, the collective interpretation of *the animals are small and big* does not result in a more specific animal-representation than the single property predication in *the animals are big*, as in both cases any animal within the set gets attributed one size-feature. Thus under this hypothesis, only intersective conjunction should elicit increased LATL activity as compared to one-property predication. However, if the computation relevant for the LATL simply reflects the attribution of features to a set but not necessarily to the same members of the set, then both intersective and collective interpretations should engage the LATL. This latter hypothesis would attribute a more general combinatory role for the LATL and indeed a finding where the LATL shows sensitivity to both intersective and collective conjunction would be compatible with many functional hypotheses, including syntactic and semantic composition in a very general sense.

The goal of Experiment 4 was to distinguish between these two hypotheses with a manipulation where participants judged the truthfulness of sentences involving conjunctions that are interpreted either collectively or intersectively. We contrasted sentences that are syntactically identical, but differed with respect to the conceptual combination of features that is required. Specifically, we contrasted plural sentences of the form *the X are P1 and P2,* where *P1* and *P2* either allowed intersection (a color and a size adjective) or strictly disallowed it (2 size adjectives that are antonyms). The predicates used in the experiment were all adjectives since, unlike verbs, adjectives more easily allow the construction of strictly incompatible pairs (e.g. *big and small*) and hence exclude the possibility of conceptual combination of those two predicates (see Chapter 3 subsection 3.7.1 for further discussion). Participants first viewed displays of objects and then indicated whether the subsequent sentence was a truthful description of the display. MEG analysis targeted the last adjective of the sentences, which was either the second adjective of the conjunction conditions or the sole adjective of our single-adjective baseline condition (Figure 1). It is important to note that our one-adjective baseline condition was itself a sentence, i.e. not noncombinatory like e.g. the one-word controls in several prior MEG studies on this topic (e.g. Bemis & Pylkkänen, 2011; 2013a, b). Thus our control condition was itself predicted to engage combinatory processing in the LATL to some extent, though, by hypothesis, less than the presentation of two adjectives. This consequently sets up the expectation that the one vs. two adjective conditions could diverge from each other more subtly in the LATL than in the prior studies where the baseline was entirely noncombinatory.

In addition to the LATL, we included as a region of interest (ROI) its right hemisphere homologue, as anterior temporal lobe (ATL) effects have often been bilateral (Mazoyer et al., 1993; Stowe et al., 1998; Friederici et al., 2000; Humphries et al., 2001; Bemis & Pylkkänen, 2011; Leffel et al., 2014), with significant debate addressing possible functional divisions between the two hemispheres (e.g. Humphries et al., 2005; Rice, Ralph, & Hoffman, 2015). Thus as an additional goal to our study we aimed to test the extent to which the pattern of ATL effects observed in the left hemisphere was mirrored in the right.

4.2 Experiment 4: Adjective conjunction and the LATL

Participants

A total of fifteen right-handed native speakers of English participated in the study (13 female; age M = 23.5). All participants were non-colorblind and had normal or corrected-to-normal vision. Before participation, they all signed an informed consent form.

Experimental design and stimuli

Stimuli consisted of visual displays followed by simple sentences which were judged for veridicality as descriptors of the display. Though somewhat artificial, our aim was to mimic naturalistic conditions where linguistic expressions are interpreted in a given context, here provided by the display. Test sentences contained adjective conjunctions, with the adjectives either allowing an intersective interpretation or forcing a collective one. The conjunction contexts were also compared to a baseline condition in which sentences contained only one adjective. The three conditions and the structure of the trials are shown in Figure 1. The target of MEG analysis was always the last (or sole) adjective of the sentence.

The intersective conjunction condition consisted of sentences containing a plural subject and two conjoined intersective adjectives: a color and an adjective describing size (e.g. *The hearts are green and big*) whereas the collective conjunction condition consisted of sentences with two conjoined adjectives that force a collective reading: two contradictory size adjectives (e.g. *The hearts are small and big*). The condition without conjunction consisted of sentences with only one adjective. Each sentence contained a period at the end, which appeared on the screen together with the final (or sole) adjective (see Figure 1). In order to measure whether participants indeed interpreted sentences in the intersective condition intersectively, and those in the collective condition non-intersectively, participants were asked to indicate for each sentence whether it was true or false for a colored image of four shapes that preceded it. Since sentences contained gradable adjectives such as *big* or *small*, participants were familiarized with the relative sizes of the figures before taking part in the actual experiment (see Procedure section below). Following our predictions, to count as a “matching trial” in the intersective condition, both adjectives (color and size) had to apply to each individual shape in the image. To count as a “matching trial” in the collective condition, one adjective had to apply to two of the individual shapes while the other applied to the remaining two shapes. In the trials without conjunction, only the shape of the objects was relevant for the decision.

Each trial contained a context image, a test sentence containing either one or two adjectives (depending on condition), and an image depicting two fingers plus “yes” and “no” buttons (Figure 1) prompting the participant to perform a task (to avoid quick button presses in response to our critical and sentence-final word). A total of six size adjectives (*big, small, thin, thick, long, short*) and six color adjectives were used (*red, blue, green, white, black, pink*) in the conjunctions. The critical sentence-final adjective was kept constant across conditions as shown in Figure 1. Collective conjunction trials were formed by replacing the first conjunct of the intersective conjunction trial (e.g. *green and big*) with the antonym of the second conjunct (yielding *small and big)*. The collective readings could only be forced with combinations of two contradictory size adjectives and thus all collective trials involved two size adjectives (a conjunction of two color adjectives can always in principle be interpreted intersectively, i.e. *pink and green*, for example, can be taken to mean ‘partly pink and partly green’). Trials without conjunction were created by removing the first conjunct (yielding *big*); since the last word of the sentence occurred with a period, participants knew to not expect a second conjunct on these trials. Predictability of the second conjunct in the collective conjunction condition was controlled for by adding a filler item to each triplet that contained the two intersective adjectives in reversed order (e.g. *big and green*).

For the noun position, twenty one-syllable nouns were used (*bag, boat, bow, cane, car, cross, cup, glass, hand, heart, house, key, lamp, leaf, lock, note, plane, shoe, star, tree*). Target images each consisted of four shapes, each unambiguously depicting one of the noun concepts, but manipulated in color and/or physical dimension.

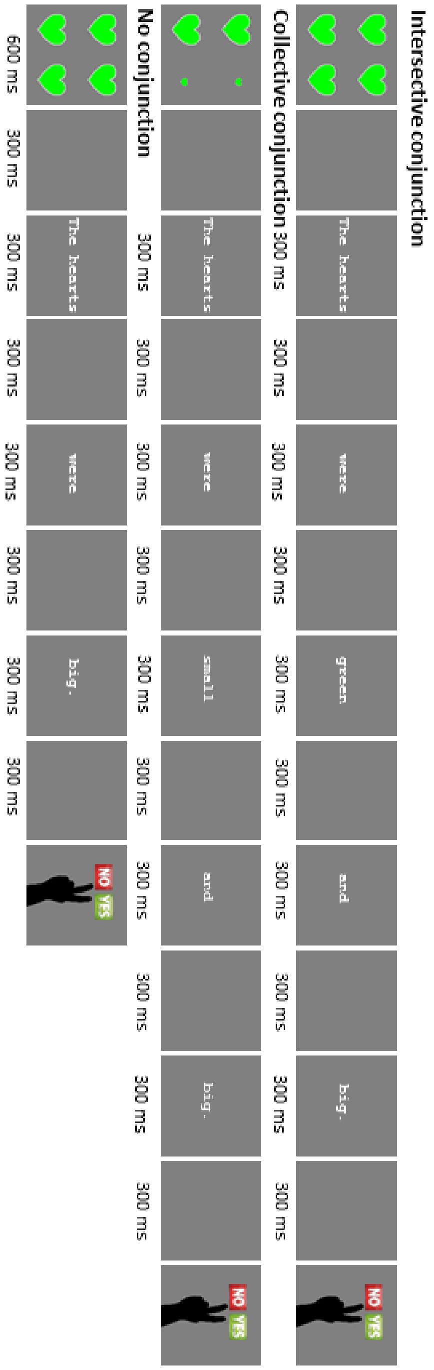

Figure 1: Experimental design and trial structure.

Participants each viewed 384 trials in 12 stimulus blocks: 96 of each of the three trial types plus 96 filler items with adjectives in reversed order. In half of the trials for each condition, the target image matched the sentence. In the conjunction conditions, half of the mismatch trials could be judged as a mismatch already on the first conjunct while the other half could only be judged as a mismatch on the second conjunct. Predictability of the final (or sole) adjective in match trials was equal across all conditions (namely 100%), assuming participants generally predict for a sentence that matches the context picture.

In each condition, each of the 20 nouns was used six times, three times in matching trials and three times in mismatching trials. The order of trials within stimulus blocks was randomized and constructed separately for each participant. Only the match trials were included in the analysis.

Procedure

Prior to taking part in the experiment, participants received written instructions on the task and saw examples of the types of images that they would see in the experiment. They then performed a practice session outside of the MEG room, in order to become familiar with the task as well as the relative sizes of the images. This session, unlike the actual experiment, provided feedback in order to ensure participants' comprehension of the task. All practice trials were distinct from experimental trials.

Before starting the MEG recording, participants' head shapes were digitized using a Polhemus Fastscan three-dimensional laser (Polhemus Inc, Vermont, USA) in order to determine the location of five marker coils that were placed across participants' heads, whose positions were measured before and after the recording. During the analysis, the digitized head shapes were used to constrain the source localization around a participant's head with respect to the MEG sensors.

During the experiment, participants lay down in a dimly lit magnetically-shielded room. They performed the task in 12 separate blocks, with the option of taking breaks between blocks. The order of blocks was randomized for each participant.

MEG data were collected using a whole-head 157-channel axial gradiometer system (Kanazawa Institute of Technology, Tokyo, Japan) with a 1000 Hz sampling rate and a low-pass filter at 200 Hz and a notch filter at 60 Hz. The screen on which the stimuli were presented was situated approximately 50 cm from the participant's eyes. During each trial, the target image was presented for 600 ms. After that, words were presented one by one for 300 ms each, and were each followed by a 300 ms blank screen. An image prompting the participant to make a decision appeared at the end of each trial and remained onscreen until the participant made a decision (Figure 1).

Pre-processing of MEG data

The raw continuously recorded MEG data were noise-reduced using the Continuously Adjusted Least Squares Method (CALM; Adachi et al., 2001). Due to urban noise, the data were then high-pass filtered using a Hamming window with a cut-off frequency of 1 Hz and a filter width of 2 s. Artifacts were manually removed from trials (from 200 ms before onset of the final adjective to 1000 ms after onset) with visual inspection by rejecting those with amplitudes exceeding +/- 3000 fT or containing eyeblinks.

The data were then averaged for each condition, from 200 ms pre stimulus to 1000 ms post stimulus (i.e. the sentence final adjective), with averages low-pass filtered at 40 Hz and baseline corrected from -100 to 0 ms prior to source analysis. Separate distributed L2 minimum norm source estimates were created for each condition average (distinguishing between match and mismatch items) for each subject, using BESA (version 6; MEGIS Software, GmbH, Gräfelfing, Germany). Next, we performed region-of-interest (ROI) analyses on the averaged data (match items only), focusing on the left and right anterior temporal lobes during the processing of the final adjective in the sentence. Secondly, we also performed a whole-brain source analysis to ensure that activity seen in the ROI analyses indeed reflected activity in the LATL or RATL and exclude the possibility of spill-over effects from adjacent areas.

In addition to analysing the activity elicited by the sentence-final adjective in all conditions, for the conjunctive conditions we also analysed the activity elicited by the first conjunct and the word *and*, in order to provide a more comprehensive processing profile of the entire conjunction and to determine whether differences between the two types of target adjectives might already be present earlier in the conjunction. Since very long epochs result in more lost trials due to artifacts, we created separate averages capturing the 'first adjective + *and*' region, starting at 200 ms before the onset of the first adjective and lasting till 1000 ms after (i.e. till 400ms after the onset of *and* or 200 ms before the onset of the second adjective). The analysis parameters for these pre-target averages were identical to those of the sentence-final adjectives.

Analysis

Behavioral responses

For each condition, we defined a 'correct' answer based on our predictions: a "yes" response in match trials. For intersective conditions, that means a "yes" response in trials where both adjectives (color and size) applied to each individual shape in the image. For collective conditions, that means a "yes" response in trials where one adjective applied to two of the individual shapes while the other applied to the remaining shapes. Finally, for no-conjunction conditions, that means a "yes" response in trials where the shape adjective applied to each individual shape in the image.

Accuracy and reaction times (RTs) of participants' behavioural responses were analyzed with repeated measures ANOVAs with Condition as the within-subjects factor (3 levels: intersective, collective, no-conjunction).

ROI analysis

The ROI analyses focused on the left anterior temporal lobe (LATL), specifically Brodmann area (BA) 38, which in prior studies has been the most consistent locus of conceptually modulated LATL composition effects (Zhang & Pylkkänen, 2015; Del Prato & Pylkkänen, 2014; Blanco-Elorrieta & Pylkkänen, 2016). The right anterior temporal lobe (RATL, BA 38) was treated as an additional ROI, as motivated above. The Talairach daemon was used to automatically assign Brodmann area labels to the 713 sources on the smooth BESA cortex (Lancaster et al., 1997; 2000) and the sources within left BA 38 were averaged together to yield a single activity time course for this region and similarly for the right BA 38. The analyses were time-locked to the presentation of the final adjective in the sentence.

Differences in ROI activity were assessed with cluster-based permutation t-tests (Maris & Oostenveld, 2007) comparing the conjunction conditions to each other as well as to the non-conjunctive baseline. Analyses were conducted at 100 – 300 ms after presentation of the target adjective, as previous studies have indicated that composition-related activity occurs at 200 – 250 ms (Bemis & Pylkkänen 2011; 2013a; 2013b). The thresholds for initial cluster selection were also adopted from prior MEG studies on the LATL (Bemis & Pylkkänen 2011; 2013a; 2013b; Westerlund & Pylkkänen 2014; Del Prato & Pylkkänen (2014): specifically, all waveform separations lasting at least 10 time samples at $p = .3$ were first grouped into clusters and of these, the one with the largest summed t-value within the analysis interval (100 – 300 ms) was entered into 10,000 permutations, yielding a corrected *p*-value (alpha = .05) that represented the proportion of times that the tests on the random partitions yielded a test statistic higher than the actual observed statistic. Finally, the *p*-values across our two ROIs were FDR corrected (Genovese et al. 2002) with a criterion value of 0.05.

In the pre-target region, this same analysis was performed at 100 – 300 ms after the onset of the first adjective of the conjunctive conditions, aimed at assessing the combinatory effect reported in prior literature. Since the general goal of the pre-target analysis was to reveal any baseline issues that might affect the interpretation of the results on the target adjective, we intentionally “double-dipped” and ran additional post-hoc permutation tests on any waveform difference that had the appearance of a possible effect. This resulted in two additional analyses at 400 – 450 ms and at 700 – 1000 ms after the onset of the first adjective. As reported below, none of these analyses revealed reliable differences.

Whole-brain analysis

In order to ensure that the ROI analyses in fact reflected activity localized to the LATL or RATL (as opposed to spill-over activity from neighboring regions), liberally thresholded uncorrected whole-brain contrasts were generated for each pair-wise comparison in order to visualize the spatial extent of the effects revealed in the ROI analysis. We performed paired t-tests in each source on the smooth BESA cortex for a time window of 0 - 700 ms from the onset of the sentence-final adjective, and plotted activity on the standard BESA brain if a source showed an effect at the $p < 0.05$ level and was surrounded by at least 2 spatial and 2 temporal neighbors that were also reliable at the 0.05 level.

Results

Behavioral data

Accuracy on average was high, with at least 90% correct responses per condition (Collect M = 94%, SD = 6%; Intersect M = 92%, SD = 7%; NoConj M = 90%, SD = 5%). Repeated measures ANOVAs were performed on both accuracy and RT data with Condition as the within-subjects factor (3 levels: intersective, collective, no-conjunction). The ANOVA on accuracy showed a main effect of Condition ($F(2, 28) = 8.99$, $p < .01$), with planned pairwise comparisons revealing that only the collective and the no-conjunction conditions significantly differed from each other ($p < .001$), with higher accuracy for the collective condition. Similarly, we found a main effect of Condition for reaction time (F (2, 28) = 24.02, $p < .001$). Here, pairwise comparisons revealed two significant differences: reaction times were longer in the no-conjunction condition compared to the collective condition ($p < .01$; NoConj M = 784 ms, SD = 28 ms; Collect M = 633 ms, SD = 26 ms) and compared to the intersective condition ($p < .001$; Intersect M = 600 ms, SD = 25 ms). These results are surprising, since the task in the no-conjunction condition required participants to check only one feature compared to two features in the other two conditions. This may be an effect reflecting surprisal, since throughout the whole experiment the no-conjunction cases comprised only a quarter of the stimulus material. Thus crucially, any increase in neural activity for the collective or intersective conditions compared to the no-conjunction condition should not reflect increased effort, since subjects were on average more accurate and faster in responding to the two-adjective conditions. Moreover, the collective and intersective conjunction conditions did not differ in accuracy or reaction time from each other. Thus again, any effects between these two conditions are unlikely to be due to differences in general effort required in the task.

ROI data: Sentence final adjective

The cluster based permutation t-tests revealed significant clusters of time points both within the LATL and the RATL, although the result patterns in the two regions were very different (Figure 2).

In the LATL, we found a significant cluster of greater activity in the intersective condition compared to the collective condition from 162 to 275 ms ($p = 0.0119$; average ± SEM: collect, 12.70 ± 1.57 nAm; intersect, 16.49 ± 1.91 nAm). When comparing the intersective to the no-conjunction condition, a cluster of increased activity in the intersective condition was found in the LATL from 181 to 232 ms, though it did not survive correction for multiple comparisons ($p = .12$; intersect, 16.30 ± 2.59 nAm; no-conj, 12.73 ± 1.90 nAm). The comparison between collective and no-conjunction conditions showed no evidence for an increase for conjunction. Instead the data trended towards the opposite pattern, with a non-significant cluster (p = .61) of increased activity for the no-conjunction condition (13.33 ± 2.49 nAm) over the collective condition (9.57 ± 1.52 nAm) occurring at 115 - 142 ms. Thus our LATL results strongly indicated a contrast in the LATL's involvement in intersective vs. collective conjunctions, with intersective but not collective interpretations increasing LATL amplitudes.

Within the RATL, the comparison between the intersective and collective conditions revealed the opposite pattern from the one observed in the LATL, with increased amplitudes for collectives over intersectives. The duration of the observed effect lasted until the end of our analysis interval and therefore we lengthened the interval to 100 – 450 ms in order to capture the entire cluster, which onset at 274 and lasted until 422 ms (collective: 21.79 ± 3.20 nAm; intersective: 15.01 ± 1.94 nAm). However, although sustained, the cluster was only near-significant after multiple comparisons correction (p = .0789). Consistent with this profile, a highly significant increase was observed for collectives when compared to the no-conjunction condition (p = 0.0184, collect 17.28 ± 2.18 nAm; no-conj, 10.90 ± 1.25 nAm). This effect also lasted through the entire analysis window (100 – 300 ms) and when the extent of the cluster was explored by lengthening the window, the cluster persisted from 0 to 516 ms (p = 0.0092, collect 17.32 ± 2.05 nAm; no-conj, 10.69 ± 1.05 nAm; analysis window: 0 – 600 ms). Finally, similarly to the pattern in the LATL, the intersective condition elicited larger amplitudes than the no-conjunction condition though this effect remained marginal (p = 0.09 for 164–229 ms, intersect, 15.54 ± 2.14 nAm; no-conj, 10.98 ± 1.36 nAm). In sum, RATL amplitudes were robustly increased for the collective interpretations and weakly so for the intersective ones.

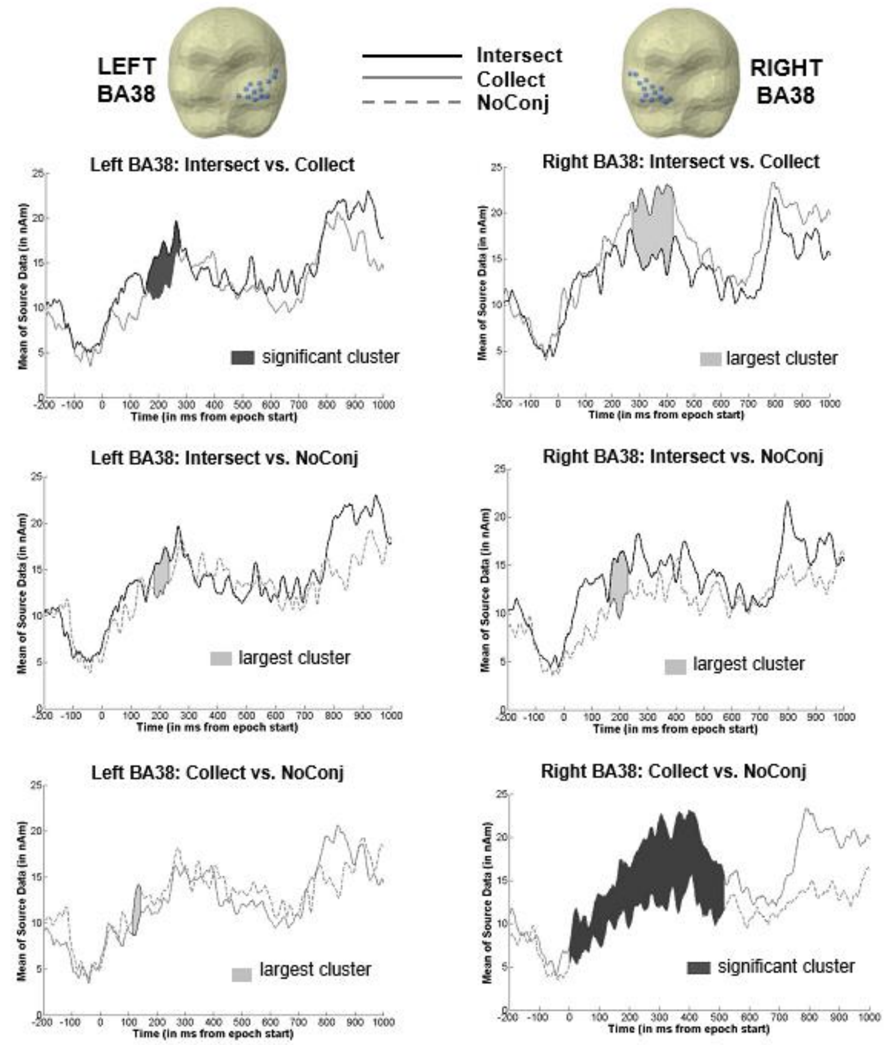

Figure 2: (A) ROI results for left and right BA 38 during the processing of the sentence-final adjective. The dark shading indicates significant clusters ($p < .05$ corrected) and the light shading clusters that did not survive correction for multiple comparisons.

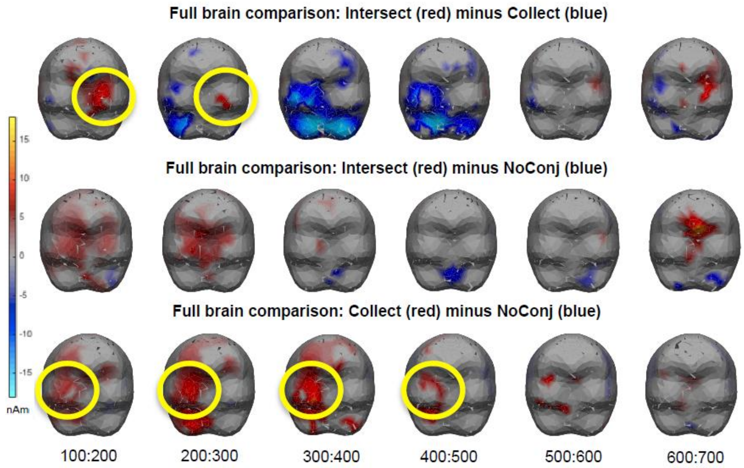

Figure 2: (B) Liberally thresholded uncorrected full brain contrasts for the same pair-wise comparisons (see Analysis), calculated in order to visualize the spatial extent of the effects obtained in the ROI analysis (in yellow circles). The effects are centered on the temporal pole, with potential spread to adjacent dorsal and ventral (though not lateral) cortex.

ROI data: First conjunct and "and"

For the conjunctive conditions, the time course of LATL and RATL activity was also assessed in the pre-target region, covering the first adjective and the conjunctive *and*, as shown in Figure 3. We searched for reliable clusters in a pre-defined time-window of 100-300 ms, aimed at targeting the combinatory effect reported in prior literature, as well as in two time-windows defined on the basis of the visual appearance of the waveforms; recall that our aim was to reveal any possible baseline issues that might affect the interpretation of our main result. The latter two analyses focussed on possible LATL increases for collectives over intersectives at ~400 ms and for intersectives over collectives at ~800 ms (top left panel of Figure 3). None of these analyses yielded reliable differences between the conditions, despite being specifically tailored to reveal such effects (the lowest uncorrected *p*-value from the tests was 0.3). Thus this analysis suggests that the differences at the sentence-final adjective in fact reflect computations occurring specifically at this adjective.

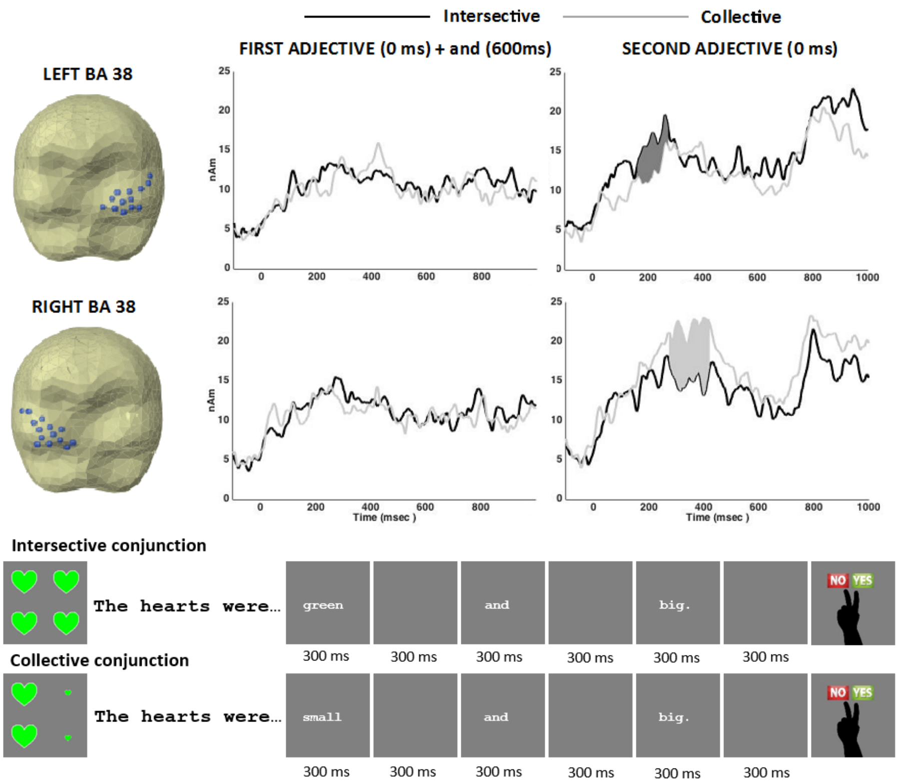

Figure 3: Time course of left and right BA 38 activity during the entire conjunction. Activity during the presentation of the first adjective and the conjunctive *and* is plotted on the left and activity elicited by the second, target adjective is plotted on the right. The analysis of the first adjective and the conjunctive *and* revealed no reliable effects, suggesting that the ATL relevant computational distinction between the two conditions only takes place at the second adjective.

Whole brain analysis

Paired t-tests over the whole brain were conducted for each pairwise comparison (Figure 2 (B)). The contrast between intersectives and collectives showed a clear activity increase for intersectives in the left temporal pole between 100 and 300 ms, consistent with the ROI analysis. Other than this, there were no obvious increases for intersectives over collectives until much later in the time course (around 600 ms). The collectives, however, elicited a robust activity increase over intersectives in the right hemisphere at 300 – 500 ms, centered on the right temporal pole as captured by the ROI analysis but also spreading posteriorly along the ventral cortical surface.

The comparison between intersectives and no-conjunction controls showed weak and somewhat diffuse increases for intersectives along both temporal poles, again consistent with the ROI analysis in which both temporal poles trended towards larger amplitudes for intersectives in this comparison.

Finally, collectives showed a robust activity increase in the right temporal pole as compared to the no-conjunction controls, as also observed in the ROI analysis. Like in the ROI analysis, left anterior temporal cortex showed no evidence for an increase for collectives in the uncorrected whole brain contrast either, further corroborating the conclusion that the LATL's computational role does not include the interpretation of collective conjunction.

Discussion

Experiment 4 aimed to move beyond the general notion of conceptual combination to more precisely characterize the computational contribution of the left anterior temporal lobe in semantic processing. Given that prior studies have already demonstrated that combinatory activity localizing in the LATL at 200-250 ms is sensitive to the conceptual specificity of the composing items (Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015), we assumed that this activity does not operate at the syntactic level but rather constructs some type of conceptual representations. More specifically, the hypothesis that the LATL reflects conceptual combination assumes that its amplitude reflects the addition of features into conceptual representations. Here we aimed to sharpen this by testing whether LATL activity reflects the number of features added to an individual or the number of features added to a set. In order to do that, we used findings on different interpretations of adjective conjunction from Experiment 2 in Chapter 3. By creating a situation where the members of a set were attributed fewer features than the set collectively, we were able to tease these two hypotheses apart: in our baseline condition, one feature was attributed to each member of the set (e.g. *the hearts were big*); in our intersective condition, two features were attributed to each member (e.g. *the hearts were green and big*), whereas in our collective condition, two features were attributed to the set but each individual only received one (e.g. *the hearts were small and big*). The main MEG analysis that we performed targeted the sentence final adjective. Our results clearly indicate that LATL amplitudes are only enhanced when the added features apply to the same individuals: distributing the features across a set did not increase this activity.

Importantly, our additional analysis of the pre-target region in the conjunctive conditions revealed no reliable differences at the first conjunct or during the processing of *and*, despite the fact that at the presentation of the first adjective within the matching trials, it was in principle possible for the subject to predict whether the conjunction would be intersective (always the case if the first adjective described all displayed objects) or collective (always the case if the first adjective did *not* describe all displayed objects). This of course crucially assumes that subjects' predictive processing is based on the anticipation of a match. The fact that our ATL effects only emerged once the conceptual content of the second adjective was revealed, suggests that either subjects did not predict as much as was possible or that these effects reflect rather bottom-up processes of conceptual combination where the actual conceptual content of the composing items matters. This latter interpretation is robustly consistent with prior findings on the LATL combinatory activity measured here, which is systemically observed even when composition is completely predictable (Bemis & Pylkkänen, 2011; 2013a) and is highly sensitive to the conceptual specificity level of the composing items (Westerlund & Pylkkänen, 2014; Zhang & Pylkkänen, 2015).

We also observed that the LATL increase elicited by intersective conjunctions is at least numerically more robust in comparison to the collective conjunction as opposed to when compared to the no-conjunction control condition, the latter contrast only eliciting a marginal effect. This was not predicted but could be due to the fact that at the single adjective in the no-conjunction condition, LATL relevant composition is likely to occur in a rather parallel way as on the second conjunct of the intersective condition. Specifically, in both positions, the property named by the currently processed adjective is attributed to all the objects described by the subject of the sentence. This parallelism could have made the Intersective vs. No-conjunction contrast a relatively weak one and in fact in many prior studies, LATL combinatory effects have been elicited against baselines that are entirely non-combinatory (e.g. Bemis & Pylkkänen, 2011; 2013a; 2013b; Brennan & Pylkkänen, 2012; Westerlund et al., 2015). This however remains speculative as with increased power, the difference between the two effects (intersective vs. no-conjunction and intersective vs. collective) could of course have flattened.

Most importantly, our findings substantially narrow down the possible hypothesis space on the computational contribution of the LATL: no account in terms of general syntactic or semantic composition can explain these results nor any general conceptualization of conceptual combination in terms of "the construction of complex concepts from simple constituents" (Baron & Osherson, 2011). Instead, the LATL does something more specific than this, potentially limited to the construction of more specific representations of individuals within a discourse model.

What counts as an "individual" for the LATL?

If this hypothesis is on the right track, future research should then target the question: what counts as a relevant type of "individual" for the purpose of LATL computation? Theories of natural language ontology can provide some guidance for how to ask this question; for example, in some theories of formal semantics, events are treated as a type of individual within the compositional system (Davidson, 1967; Parsons, 1990) whereas in others they are not (e.g. Montague, 1970; Heim & Kratzer, 1998). In the former type, so-called event-based theories, verbs are treated as properties of events and most nouns as properties of individuals, resulting in a system in which events and individuals function as similar type of variables in the combinatory system. For example, the hypothesized interpretation of *red square ball* could be paraphrased as “an individual such that it is a ball, red and square” and the interpretation of a verb phrase such as *stab Brutus in the back* would similarly proceed as “an event such that it is a stabbing, has Brutus as an affected participant, and is in the back” (Parsons, 1990). Thus in this type of theory, the direct object *Brutus* ends up as a concatenated property in the event description, similarly to modifiers such as *in the back* or *red* and *square* in the noun phrase. A relevant empirical question then is: does the construction of event descriptions engage the LATL similarly to the construction of entity descriptions, such as *red ball*? Westerlund et al. (2015) addressed this in a design targeting a possible computational distinction between modification and argument saturation, with results showing that the composition of phrases such as *eats meat* engages the LATL in a similar fashion to adjective noun combinations. Thus, if the LATL constructs representations of individuals, events also count as such individuals.

‘Hub and spoke’ model

In broad terms, our results allow an interpretation along the lines of a ‘hub and spoke’ model of conceptual knowledge, as discussed in subsection 4.1.1 (e.g. Patterson et al., 2007; Lambon Ralph et al., 2010). Many studies have identified the anterior temporal lobes as candidates for the localization of this hub, by showing that damage to these regions causes Semantic Dementia. Semantic Dementia is shown to affect categorization performance depending on the specificity with which items must be recognized: specific concepts are more affected than less specific ones (Warrington, 1975; Hodges et al., 1995; Done & Gale, 1997). In the current study, the intersective conditions required the combination of two distinguishing features such as *big* and *green,* resulting in a more specific concept (i.e. *big green hearts*) than in the collective conditions, which do not require such integration of features. Instead, in the case of for example *big and small hearts*, two higher-level concepts are distributed over entities. These concepts are more robust in the sense that they are distinguishable by several different features, resulting in a less distinctive classification. Consequently, the intersective conditions, with more specific concepts like *big green hearts* are expected to engage the semantic hub more than the collective conditions, because there are more overlapping representations to differentiate. Simply put, the hub needs to instantiate the ‘heart’ representations more exactly for *big and green hearts* (in terms of both shape and color) than for *big and small* hearts (where it’s merely necessary to find ‘heart’ representations that are sufficiently like big hearts or small hearts in order to activate the concept). Consequently, the increased LATL activity in the intersective over the collective conditions might reflect the difference in specificity of concepts that are being categorized in these conditions – following the predictions of the ‘hub and spoke’ model. This result is in line with neuroimaging evidence that shows more LATL activity for more specific image categorization (for example in case of subordinate categories like *robin*) than for basic level or domain general level categorization (for example in case of *bird* or *animal*) (Rogers et al., 2006). It is also in line with Westerlund and

Pylkkanen's (2014) MEG results on small adjective phrases, which show that LATL activity increases proportional to the change in concept specificity, as well as the findings that their paper was based on, namely that adjective-noun constructions like *blue boat* lead to more LATL activity than simple nouns like *boat* (Bemis & Pylkkänen, 2011; 2013a; 2013b).

Left vs. Right ATL

Interestingly, the RATL, which was not primarily part of our hypotheses but is often thought to be responsible for similar processes as the LATL, in fact showed a very strong opposite effect from the LATL. In the RATL, the collective conditions led to significantly more activity than the one-adjective baselines and trended towards larger amplitudes than the intersectives. This is of theoretical interest since the left and right ATLs often show similar effects (Ferstl et al., 2008; Lambon Ralph et al., 2009, 2010; Visser et al., 2010), including in MEG work coming from the same lab as the current study (Bemis & Pylkkänen, 2011; Leffel et al. 2014). The respective functional contributions of the ATLs have thus been non-trivial to characterize. However, in light of more general proposals about the distinct roles of the left and right hemispheres in language processing, our hemispheric difference is less surprising. The left hemisphere is typically thought of as the faster, more local processor whereas the right hemisphere contributes to more global computations (e.g. Beeman & Chiarello, 1998). Most relevantly, there is evidence that the right hemisphere contributes to the establishment of global coherence while the left hemisphere computes local coherence relations (St. George et al., 1999). Our results could be predicted from this hypothesis: the collective readings are in a sense locally incoherent, containing a conjunction of contradictory adjectives which only become coherent when distributed to distinct members of the set named by the subject of the sentence. In contrast, the intersective conjunctions are locally coherent as they describe two mutually compatible properties. Additional evidence for a right hemisphere role in processing antonymic relations comes from neuropsychological data on right hemisphere patients who are especially bad on tests measuring their sense of opposition (Caramazza et al., 1976; Gardner et al. 1978).

The timing profiles of the two ATL effects also fit the general idea that the left hemisphere performs faster and more local computations than the right hemisphere: in the left ATL, intersectives elicited larger amplitudes than collectives at 200-250 ms while the reverse, though statistically weaker, pattern obtained in the right ATL roughly one hundred milliseconds later (~300-400 ms). In comparison to the one-adjective baseline though, the right ATL increase began very early, in fact at the onset of the stimulus (and thus presumably even earlier). Most likely this reflects anticipatory activity reflecting a prediction for a collective reading: whenever the first adjective truthfully described only a subset of the just viewed objects, the only conjunction that could match the picture was a collective one. Thus assuming that subjects in general predicted the sentences to match the pictures, they should have been able to anticipate the presence of a collective reading. Some of this activity could, however, also be associated with general anticipation for a second adjective, since intersectives exhibited a weak trend in the same direction.

Atypical predicate combinations

Finally, going back to Chapter 3 which led to the stimuli that were used in Experiment 4, an interesting question concerns LATL activity when processing what I called “atypical” cases of predicate conjunction. In Chapter 3, I argued that any theory of predicate conjunction needs to take into account more than merely cases of conjunction that are always interpreted intersectively and those that are always interpreted non-intersectively. Experiment 3 in Chapter 3 showed that there are cases of conjunction where the predicates are not strictly incompatible, but are still acceptable in non-intersective, split situations. For example, a sentence like *The men are sitting and cooking* is often accepted in a situation in which some men are sitting and some men are cooking, despite the fact that an intersective interpretation in which each man is both sitting and cooking simultaneously is also possible. Based on such observations, I predicted that the Maximal Typicality Hypothesis (MTH) applies to sentences with predicate conjunction (see sections 2.3 and 3.4 for detailed explorations). The MTH makes predictions about sentence interpretation based on typicality information on the conjoined concepts. According to the MTH, sentences are accepted in a non-intersective, split situation to the degree that it is atypical for the two predicates to apply simultaneously to one individual (i.e. their compatibility). Experiment 4 in the current chapter employed only the two extremes on the spectrum that the MTH assumes: conjunctions that always result in intersective interpretations and conjunctions that always result in non-intersective, collective interpretations. This was done in order to clearly distinguish between cases where two features are always attributed to each member of a set and cases where two features are attributed to a set but never to each individual in that set. An interesting follow-up study would also include atypical cases of conjunction. In Experiment 3, I found a significant correlation between interpretation on the one hand and an independent compatibility measurement on the other hand, following the predictions of the MTH. Based on that finding, I predict to also find a correlation between a compatibility measurement and LATL activity. In particular, the prediction would be that the degree to which two predicates are judged as compatible correlates with the degree to which two features are attributed to each member of a set, and hence with the amplitude of LATL activity. Obtaining such a result would be additional support for the MTH and would strengthen the conclusions from Experiment 4.

4.3 Conclusion

In this work we used the findings on adjective conjunction from Chapter 3 in order to characterize the nature of the conceptual combination that the LATL is responsible for. We compared two types of conjunctions that differed in their interpretation, and hence in terms of the way they distribute two features to members of a set: either the two features are added to one representation (intersective conjunction) or they are each distributed to different members of a set (collective conjunction). By showing increased activity in the LATL for intersective conjunctions over collective conjunctions, we provide evidence in favor of the hypothesis that the LATL is responsible for the combination of features to form one coherent entity representation. The pattern observed in the RATL was the opposite, indicating that its function in terms of conceptual combination is clearly distinct from the LATL.

Chapter 5
On plurality and the effects of spatial knowledge

The first part of this dissertation reported experimental work that studied the interpretation of complex plural sentences. Chapter 2 investigated plural sentences with reciprocal expressions, and Chapters 3 and 4 focused on plural sentences with predicate conjunction. I argued that these two types of plural sentences have something in common: their logical interpretation is systematically affected by conceptual aspects of content word meaning. In reciprocal sentences, interpretation is affected by the typicality of different situations for verb concepts like *pinch* or *know*. In plural sentences with predicate conjunction, interpretation is affected by the typicality of situations for verb concept combinations like *sitting and cooking* or *sitting and reading*. In more general terms, I concluded that the truth conditions of such plural sentences should be stated relative to a specific contextual factor: lexical-conceptual knowledge.

While the dissertation has thus far focused solely on the effects of this type of context, an interesting question concerns whether there are other contextual factors that affect interpretation of plural sentences. This chapter explores one other potential candidate: spatial knowledge. Experimental results on the interpretation of reciprocal sentences reveal that situations that differ merely in the spatial configuration of the antecedent set result in different acceptability patterns, suggesting that spatial knowledge is another contextual parameter that affects the interpretation of plurals.

5.1 Flexible sentence meanings and spatial knowledge

Because experiments 1 through 4 aimed to measure specifically what the effects of lexical-conceptual knowledge are on sentence interpretation (and, in the case of Experiment 4, also on brain activity), an effort was made to keep all other contextual factors constant in the interpretation part of the experiments.

In Experiment 1, participants judged the acceptability of reciprocal sentences such as *John, Bill and George are pinching each other* in accompanying images. Items differed only in terms of the content words in the sentences (e.g. *pinching* vs *biting*). Other contextual factors were kept neutral. Sentences were presented in isolation without any prior contextual information about the individuals that the sentence referred to, and the exact same images were used for each test sentence (schematic representations of situations).

In Experiment 2, 3 and 4, participants judged the acceptability of plural sentences with predicate conjunction such as *The men are sitting and cooking* or *The animals are big and small* in accompanying images. The only difference between items was in terms of the content words in the sentences (e.g. *sitting and cooking* vs. *sitting and standing*, or *big and small* vs. *big and gray*). Again, other contextual factors were kept constant. Sentences were presented without prior contextual information. All test images in Experiment 2 and Experiment 3 depicted four entities next to each other in a row. In Experiment 4, four entities were always depicted in two rows of two entities.

Results revealed that sentence interpretation differs when content words differ. Since I made sure that items only differed in terms of lexical content, I was able to ascribe that effect to that particular manipulation. But is conceptual information on content words the only contextual factor that affects sentence interpretation? What happens when other factors in the context are manipulated instead?

I believe that sentence meaning is flexible and that lexical knowledge is only one – albeit crucial – factor that can cause the logical interpretation of sentences to change. For plural reciprocal sentences, this intuition was first spelled out by Dalrymple et al. (1998), who propose that sentence meaning is affected by linguistic and non-linguistic context in a general sense. That would imply that as soon as other contextual factors besides lexical knowledge change, a speaker's interpretation of a sentence can also change. In this chapter I focus on one other specific factor that I hypothesize to have an effect on logical sentence interpretation: spatial knowledge. Consider for example the reciprocal sentence (1), which was also mentioned by Dalrymple et al. (1998) in a slightly different formulation (p.169-170).

(1) The children are sitting alongside each other

The effect of spatial context on the logical interpretation of (1) is quite clear if we contrast two different situations that differ only in terms of the spatial arrangement of the children: one in which they are sitting on a bench and one in which they are sitting around a campfire. Assume that there are five children. If we know that they are sitting on a bench, then our interpretation of (1) is probably such that they are sitting in a line configuration, i.e. there are two children at the ends of the line who are only sitting alongside one other child. In that situation, the logical interpretation is thus such that there are four symmetric 'sit alongside' relations[35] among the five children (Figure 1A). By contrast, if the children are sitting around a campfire, then they are sitting in a circle

[35] To be more precise: I use 'four symmetric relations' for simplicity to refer to eight ordered pairs in the symmetric relation *sit alongside*, which are based on four doubletons. I use 'symmetric relations' in the same way later on in the paragraph.

configuration, making it possible for every child to sit alongside two other children (thus creating a situation with five symmetric 'sit alongside' relations among the five children (Figure 1B)). The critical question that arises then is whether (1) is true in a circle configuration with four symmetric 'sit alongside' relations as well, or whether the spatial possibility of having five symmetric relations causes the sentence to be false in situations with fewer relations? Intuitively, it seems like the latter is the case, i.e. (1) is true in a circular configuration only when there are five symmetric relations as in Figure 1B.

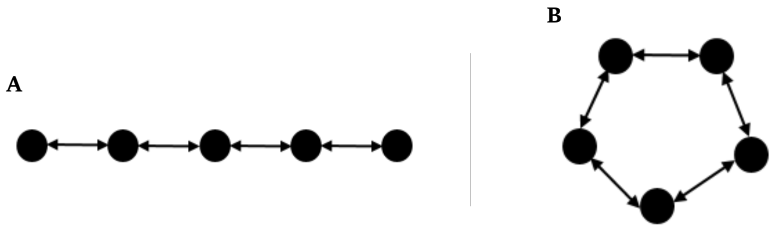

Figure 1: Line configuration with four symmetric relations (A) and circle configuration with five symmetric relations (B) for *sitting alongside each other*.

This example indicates that sentence interpretation might be affected by our spatial knowledge. When sitting on a bench, five symmetric relations are not possible because the child far left cannot sit next to the child far right. By contrast, when sitting around a campfire, five symmetric relations are very natural. In this chapter, I further explore these kinds of effects. In order to systematically study the effects of spatial knowledge, one needs to single out that particular contextual factor in an experiment – just like the experiments in this dissertation have done to study specifically the effects of conceptual knowledge.

5.2 The effects of spatial knowledge on reciprocity (with preliminary results)

The way to single out the spatial knowledge factor is to compare the acceptability of a sentence in two situations that differ merely in terms of that factor. This approach is identical to the approach in experiments 1 through 4 that study the effects of conceptual knowledge, where the only difference between items was in terms of conceptual information. Consider for example Figure 2A and 2B given the sentence *Mary, Sue and Jane are holding hands with each other*.

A **B**

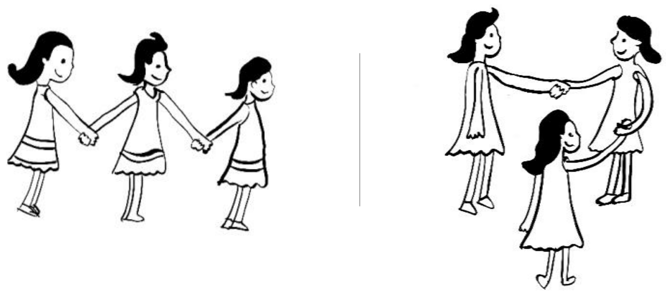

Figure 2: Three girls holding hands in a line configuration (A) and a circle configuration (B).

These two situations differ in terms of their spatial configuration. In 2A, the girls are situated in a line configuration, while in 2B the girls are situated in a circle configuration. Other factors are constant, such as the number of girls, the appearance of the girls, and the number of 'holding hands' relations among the girls. In all respects expect for the spatial setting, 2A and 2B are identical. That means that if we now measure acceptability of the same sentence in each situation, any difference in acceptability rate has to be due to that spatial setting. Preliminary results in Poortman (2011) reveal that with symmetric verbs[36] like *hold hands*, speakers accept reciprocal sentences more often in a line configuration with two symmetric relations (as in 2A) than in a circle configuration with two symmetric relations (as in 2B), despite the fact that these are identical with respect to the number of relations among the antecedent set members.[37] I hypothesize that this is because speakers first consider the three individuals in the given configuration, and then apply a mechanism that selects the maximal meaning given that configuration. If we take the configuration to be such that three girls are standing next to each other in a line, then the maximal interpretation of the sentence *Mary, Sue and Jane are holding hands with each other* is a situation such as 2A. We cannot add more relations to this configuration because according to our spatial knowledge it is simply impossible for the girl far right to act on the girl far left. By contrast, if we take the configuration to be such that the girls are standing in a circle, then the maximal interpretation of that same sentence is not situation 2B, but a situation in which the circle is closed – i.e. when there are three symmetric relations. In other words, situation 2B is not maximal given the spatial configuration of girls, because more relations can be added to the situation by closing the circle. Due to this

[36] With the term *symmetric verb* I refer to a verb for which it follows that if *A verb B*, then also *B verb A*.

[37] In Poortman (2011), I show that with non-symmetric verbs like *pinch* or *bite*, both line configurations with two relations and circle configurations with two relations are generally unacceptable. This is in line with the results on *S2* described in Experiment 1 in Chapter 2. Spatial effects of the sort described here only surface with symmetric verbs.

difference, the sentence is more acceptable in the maximal 2A than in the non-maximal 2B.

A follow-up experiment that I performed in the context of this dissertation, which focuses specifically on reciprocal sentences with symmetric verbs, reveals this effect more substantially (for all details see Appendix C). The experiment measured speakers' acceptability of reciprocal sentences in situations where spatial knowledge either allows three symmetric relations (as in 2B) vs. situations where it allows only two such relations (2A). Eleven reciprocal sentences of the form *A, B and C are P-ed to each other* (where *A, B* and *C* refer to three pieces of material and *P* is a symmetric verb) were tested on their acceptability in two different test images. One image depicted a line configuration with two symmetric relations among the materials. The other image depicted a circle configuration with two symmetric relations, while there could have been a third one – thus creating an 'incomplete' circle. In all other respects, the images were as identical as possible. An example is in Figure 3.

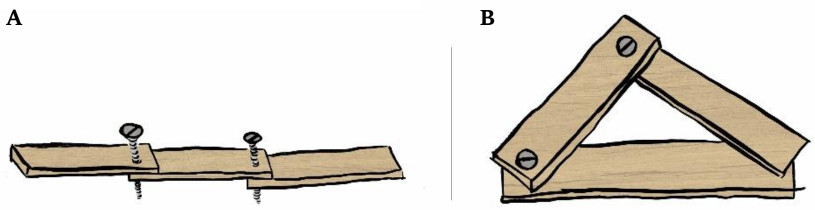

Figure 3: Test images for *The wooden planks A, B and C are screwed to each other*: line configuration (A) and circle configuration (B).

Participants were instructed to rate how well each sentence describes the accompanying image on a scale of 1 to 5 (from very badly to very well). Results revealed a significant main effect of configuration type (F (1,27) = 29.61, p < .001). The rating for reciprocal sentences in linear configurations was significantly higher than in circular configurations (M = 3.88, SD = .74 and M = 3.03, SD = .77, respectively). That means that the reciprocal sentences that were tested are more acceptable in the line configurations than in the circle configurations, even though the number of symmetric relations among the objects is the same in both situations (i.e. two) and no other contextual factors are different.

These results show that besides conceptual knowledge (as tested in Experiment 1 in Chapter 2), spatial knowledge also affects interpretation of reciprocal sentences. The knowledge that for example the leftmost and rightmost item in Figure 3A cannot be screwed together in the line configuration causes the acceptability of the reciprocal sentence to be higher than in a situation in which those two items could have been screwed together but are not (Figure 3B).

5.3 The effects of spatial knowledge on plural predicate conjunction

A logical next step is to investigate whether spatial knowledge is a contextual factor that also affects the interpretation of plural sentences with conjoined predicates, as were in the focus of Chapters 3 and 4. In the experiments in those chapters, the manipulation was always the pair of content words in the sentences. What happens if we keep those constant – thus we look at one sentence – but manipulate the context in which the sentence is evaluated in terms of its spatial configuration? Despite the fact that there are no experimental data on this issue, this subsection speculates that such effects exist, and provides suggestions for how they could be tested.

In order to investigate specifically whether spatial knowledge can affect the logical interpretation of conjunctive sentences, we would need to compare the acceptability of a sentence in two situations that differ merely in terms of spatial factors. A candidate sentence is (2), taken from Winter (2001a).

(2) The birds are flying above the house and below the cloud

Firstly, assume that we interpret the conjunction in (2) in a context where the cloud is situated right above the house (as in Figures 4A and 4B). In such a spatial context, (2) is probably only true in case all birds are flying in between the house and the cloud, as in Figure 4A. The logical interpretation of *and* is simply set intersection: each of the birds is in the intersection of the set of individuals that are above the house and the set of individuals that are below the cloud, i.e. each bird is above the house and below the cloud. Sentence (2) is likely to be judged false in Figure 4B, which would be consistent with a non-intersective analysis such that some birds are above the house and some birds are below the cloud (a ‘split situation’).

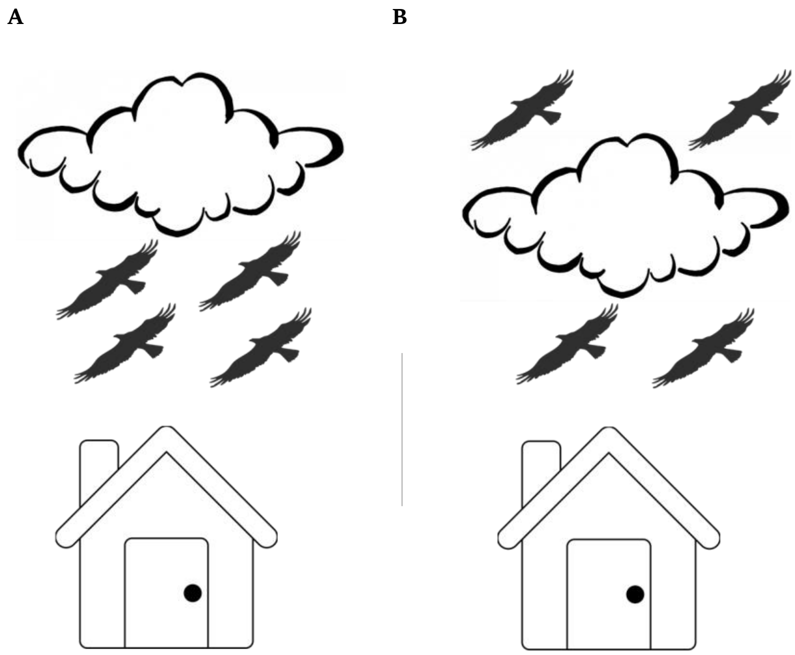

Figure 4: Two situations with a cloud right above a house

By contrast, if we are in a context where the placement of the house in relation to the cloud is different, as in Figure 5, then our logical interpretation of (2) likely changes.[38] Our spatial knowledge now does not allow an intersective interpretation of *and*, since it cannot be the case that each bird is both above the house and below the cloud (at least not in the strict sense). Hence, in such a situation we do interpret the sentence non-intersectively: the sentence is true if some birds are flying above the house and the remaining birds are flying below the cloud, i.e. in a split situation.

[38] This example was mentioned to me by Yoad Winter, who attributes it to Tim Stowell (p.c.).

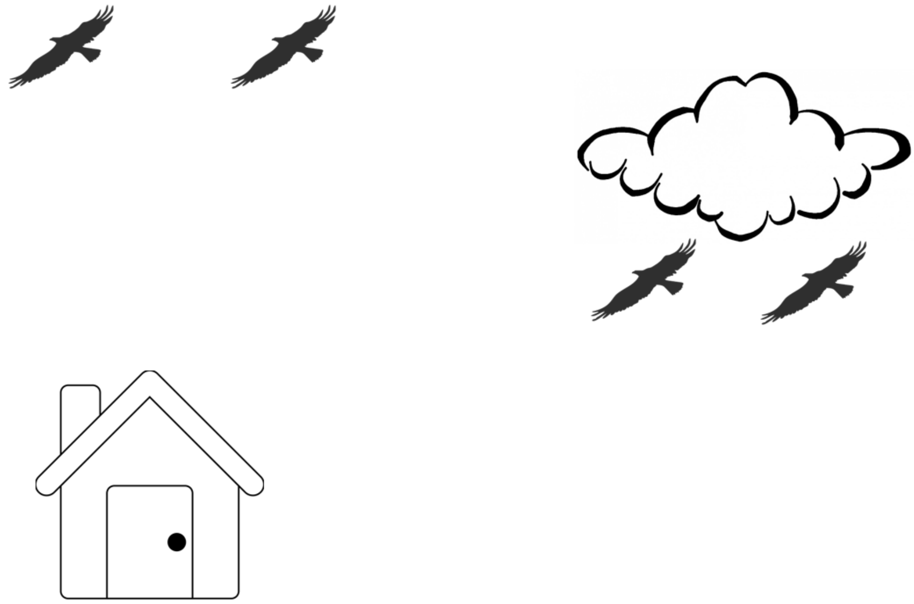

Figure 5: A situation with a cloud situated away from a house

In sum, the interpretation of sentence (2) is likely to be dependent on the spatial setting that it is evaluated in. When that setting allows an intersective interpretation of *and*, than (2) is only true in situations that are consistent with such an interpretation (Figure 4A), hence false in a split situation (Figure 4B). When the setting does not allow an intersective interpretation of *and* because our spatial knowledge prevents it, then (2) is true in a split situation that is consistent with a weaker, non-intersective interpretation (Figure 5). This effect cannot be ascribed to conceptual knowledge about the conjoined predicates (as we did in Chapter 3 and 4), but has to be purely due to the spatial differences between 4A/B and 5. This is a similar phenomenon to sentence (1) being acceptable in both a line configuration and a circle configuration – depending on the spatial situation. An experiment that tests this effect systematically would need to measure speakers' acceptability of sentences with predicate conjunction in situations where spatial knowledge either allows an intersective interpretation vs. situations where it does not - similar to the experiment on reciprocals that was reported in the previous subsection.

5.4 Summary and conclusion

The main part of this dissertation has been about the effects of lexical knowledge on the interpretation of complex plural sentences. This short discussion chapter aimed to convey that spatial knowledge is an additional contextual parameter that needs to be taken into account in a theory of plurals. The experimental work on reciprocal sentences with symmetric verbs indicates that spatial information indeed seems to affect the logical interpretation of a sentence, and there is good reason to believe that similar effects exist with plural sentences containing predicate conjunction. Further experimental research on these sentences will help specify spatial effects more precisely.

More generally, any future experimental work on plural sentences will need to control for spatial effects, as was done in the experimental work in this dissertation. As mentioned, when measuring sentence interpretation in Experiments 1 through 4, an effort was made to keep all spatial factors constant, so that any difference in interpretation had to be due to the manipulation of conceptual knowledge.

Chapter 6
On plurality and concept composition

The Maximal Typicality Hypothesis that was introduced in this dissertation is a meaning principle that operates whenever there is a conflict between typicality and maximality within a plural sentence. For example, in *The men are pinching each other*, there is a conflict between what is typical for the verb concept *pinch* and what is maximal for the logical operator *each other*. A similar conflict occurs with predicate conjunction in plural sentences such as *The men are sitting and cooking*. In such a sentence, there is a conflict between what is typical for the verb concept combination *sitting and cooking* and the strongest (or "maximal") interpretation of *distributivity* combined with the connective *and*. Interestingly, similar conflicts also occur in a different domain, which has been studied more extensively: nominal modification constructions such as *red hair* or *pet fish*. The expression *red hair* often refers to the type of hair called 'ginger', which is not the most typical for the concept *red*, nor the most typical for the concept *hair*. Similarly, *pet fish* most commonly refers to a type of fish that is not the most typical fish, nor the most typical pet. This chapter discusses how results on lexical-conceptual effects within plural sentences relate to concept composition effects with nominal modifiers.

6.1 Concept composition in modification constructions: typicality and categorization

According to the well-known 'compositionality principle', the meaning of a complex expression is derived from the meaning of its parts and the way they are combined (e.g. Janssen 1997). In nominal modification constructions, probably the simplest way of following the compositionality principle is using Boolean conjunction. For example, the meaning of the expression *leather jacket* follows from the meaning of *leather* and the meaning of *jacket*, plus Boolean conjunction, such that a leather jacket is something that is categorized as *leather* and categorized as a *jacket*. However, there are some challenges for this straightforward logical rule in concept composition.

Firstly, the question arises how the prototype of a composed concept is constructed. Can we predict the typicality of an instance for the composed concept AB from its typicality for the conjoined categories A and B? Osherson and Smith (1981) point out examples where the typicality of an object for AB does not follow straightforwardly from the typicality of that object for A and B individually, because there is a conflict in typicality between A and B. The most well-known example is *pet fish*, constructed by combining the individual categories *pet* and *fish*. A typical exemplar of a *pet* is soft and can be stroked and petted (e.g. dog, cat, rabbit), while a typical exemplar of a *fish* is one that is caught in nets and served with fries (e.g. cod, trout). The most typical exemplar of the composed concept *pet fish*, however, is a guppy, which has none of those properties. That implies that a guppy is a more typical instance of the composed concept *pet fish* than of either of the concepts *pet* or *fish* in isolation. This is hard to explain if we assume that the typicality of an object for a complex concept follows from its typicality with respect to the constituent concepts. Clearly, the typicality of a guppy as a *pet fish* does not follow from its typicality as a *pet* nor from its typicality as a *fish*. Such effects have become known as 'guppy effects'.[39]

Another question concerns categorization within a composed concept AB. While Osherson and Smith take guppy effects in typicality to be a challenge for the study of concepts and how typicality values are constructed, they do not take them to be a challenge for categorization. The authors argue that categorization follows classical logical rules. According to Osherson and Smith then, even when typicality effects with complex concepts are not easily understood, categorization using such concepts is nevertheless straightforward. For example: despite the challenging prototypical exemplar of *pet fish*, an entity is categorized as a *pet fish* simply if it is independently categorized as a *pet* and as a *fish*. Hampton (2007) argues against this proposal, by showing that 'non-compositional', guppy-type effects also occur in categorization. That means that just like typicality, category membership within a composed concept AB is also not always easily predicted from its constituent concepts A and B. Sometimes membership in a conjunctive concept is not a simple logical combination of membership in each of the conjoined individual categories. For example, Hampton (1988) shows that speakers sometimes categorize chess as an instance of the complex concept *sport which is also a game*, but not as an instance of the concept *sport*. This indicates that the logical conjunctive rule fails in categorization as well: sometimes an object is categorized as an instance of the concept AB, but not as an instance of A (or B) in isolation. Hampton calls these effects 'overextension effects' (Hampton, 1988).

An important factor affecting category membership of complex concepts is typicality conflicts between the two conjoined individual concepts. Lee (to appear) finds a clear effect of typicality conflicts on categorization in modification constructions, by contrasting cases where there is a typicality conflict between the conjoined categories with cases where there is no such conflict. An example of such a contrast is *red hair* vs.

[39] While I use the key example *pet fish* to explain the phenomenon of 'guppy effects', Storms et al. (1998) claim that they could not experimentally verify such a guppy effect with this particular example. However, guppy effects were attested for other examples in many experimental studies (including Storms et al.'s), so I do not take this to be an issue for the arguments in this chapter.

red car. The expression *red hair* contains a conflict because instances that are typical for the concept *red* are atypical for the concept *hair* (i.e. focal red hair is not typical *hair*), and vice versa instances that are typical for the concept *hair* are atypical for the concept *red* (i.e. 'ginger' hair is not a typical *red* hue). Such conflicts do not arise for *red car*. Participants in Lee's study very often categorize 'ginger' hair as *red hair* (92% of participants), while they hardly ever categorize a car with that same 'ginger' hue as a *red car* (17%). This indicates that the standard of redness in a modification construction is affected by conceptual meaning aspects of the noun that *red* combines with. When there is a conflict in typicality between the modifier *red* and the noun *hair*, that conflict is resolved by adjusting the standard of redness from a typical hue to a less typical one. A similar phenomenon is described in Kamp and Partee (1995), who refer to it as "context-sensitive recalibrations of predicates" (p. 164). How does this relate to conclusions from Chapter 2 and 3 that the interpretations of *and* and *each other* are affected by conceptual meaning aspects of the predicates that they combine with?

6.2 Concept composition in a plural verb phrase

Winter (to appear) points out that many of the plural sentences that were investigated in Chapters 2 and 3 can also be said to contain a conflict between two (or more) concepts. Consider sentences (1) and (2) (reproduced from Chapter 2 and 3, respectively).

(1) The men are pinching each other
(2) The men are sitting and cooking

If we assume that a VP like *pinch each other* is composed of the vague verb concept *pinch* and the vague logical concept RECIPROCITY, and a VP like *sitting and cooking* is composed of the vague verb concepts *sitting*, *cooking* and the vague logical concept DISTRIBUTIVITY combined with *and*, then the phenomena that were studied in Chapter 2 and 3 can be phrased in similar terms to the phenomenon of modification constructions such as *red hair*. Just like with *red hair*, concepts with conflicting typicality preferences combine to form a complex expression in (1) and (2). I have introduced the MTH as the mechanism that resolves such conflicts for the two plural constructions that were studied in this dissertation. Winter gives a uniform account of all three constructions, i.e. modification constructions, plural reciprocal constructions and plural conjunctive constructions. To underline the parallelism that he proposes, let us first describe the conflicts in (1) and (2) in these terms in more detail.

Pinching each other

With the concept *pinch*, Experiment 1 in Chapter 2 suggests that a situation with one patient per agent (e.g. *S2*, *S3*) is a more typical instance than a situation with two simultaneous patients per agent (e.g. *S6*). With the concept of RECIPROCITY, we did not measure which instance is most typical. However, it is reasonable to assume that the most typical instance is the maximal one: Strong Reciprocity. That means that in a context with three individuals, a situation where every individual pinches both other individuals (*S6*) is more typical for RECIPROCITY than a situation with fewer relations among the individuals. When we now combine the concept *pinch* with the concept RECIPROCITY as in (1) (assuming a context of three individuals), there is a conflict in typicality. While *pinch* prefers each agent to act on only one patient (*S2*, *S3*), RECIPROCITY prefers each agent to act on both other patients (*S6*). In sum:

concepts	typicality
C1: *pinch*	*S2* = *S3* > *S6*
C2: RECIPROCITY	*S6* > *S3* > *S2*

Sitting and cooking

Experiment 3 in Chapter 3 reveals that with the concept pair *sitting* and *cooking*, a situation where an individual does not both sit and cook simultaneously is a more typical instance than a situation where an individual sits and cooks simultaneously. With the (covert) concept of DISTRIBUTIVITY, we assume that the most typical instance is universal quantification. That means that when DISTRIBUTIVITY combines with Boolean *and* in a context with four individuals, a situation where every individual is sitting and every individual is cooking is more typical than a situation with fewer actions among the individuals. When we now combine the concepts *sit*, *cook* and DISTRIBUTIVITY + *and* as in (2) (assuming a context of four individuals), there is a conflict in typicality. While the combination *sit* and *cook* prefers situations where no individual sits and cooks at the same time (e.g. split situations, incomplete split situations), DISTRIBUTIVITY + *and* prefers each individual to do both simultaneously (joint situation). In sum:

concepts	typicality
C1: *sit and cook*	incomplete split situation = split situation > joint situation
C2: DISTRIBUTIVITY + *and*	joint situation > split situation > incomplete split situation

In Chapter 2 and 3, the MTH was put forward as the principle that resolves these types of conflicts within plural sentences. The interpretation of complex expressions like *pinching each other* or *sitting and cooking* is a 'compromise' between the conflicting typicality preferences of its parts. This compromise is what we called the 'core situation' for a sentence, as selected by the MTH. The core situation is maximally typical for a sentence, taking into account the typicality effects invoked by the separate concepts C1 and C2 that it contains. Specifically, the core situation for a sentence is the situation that attains maximal typicality for C2 (the logical concept) among the situations that attain maximal typicality for C1 (the verb concept(s)). This allowed us to explain experimental data on sentences that previous accounts failed to account for, namely those where typicality effects were in conflict:

> The core situation that the MTH selects for (1) is *S3*, i.e. every man pinches one other man and is being pinched by one other man. This is the optimal compromise between the conflicting typicality preferences because *S3* is the maximally typical situation for the concept RECIPROCITY among the situations that are most typical for the concept *pinch*. Therefore, (1) is accepted in *S3* (as tested in Experiment 1).
>
> The core situation that is selected for (2) is the split situation, i.e. two men are sitting and two men are cooking. This is the optimal compromise because it is the maximally typical situation for the concept DISTRIBUTIVITY among the situations that are most typical for the concept pair *sitting* and *cooking*. Therefore, (2) is accepted in the split situation (as tested in Experiment 3).

6.3 The MTH in modification constructions

After pointing out the parallelism between conflicting typicality for concepts in modification constructions and in plural VPs, let us now discuss the possibility of a common solution for conflict resolution in these constructions. Winter (to appear) proposes such a general principle. He argues that for all these examples, there exists critical typicality points: points that attain maximal typicality for the gradable concept (e.g. *red*, RECIPROCITY, DISTRIBUTIVITY) among the points that attain maximal typicality for the head concept (*hair*, *pinch*, or the bi-concept *sit&cook*). Winter uses these critical typicality points to help predict acceptability behavior, arguing that "critical typicality points are an important factor that boost acceptability judgments for complex expressions" (Winter, to appear). What I have referred to as the 'core situation' for a plural sentence is similar to Winter's critical typicality point for such sentences. Let us have a look what happens when we apply either Winter's uniform proposal or something like the MTH to modification constructions like *red hair* or *pet fish* (for which we would have to rephrase the MTH so that it selects a 'core exemplar'[40] instead of a 'core situation'). The critical typicality point or core exemplar is the one that attains

[40] The idea of a maximally typical exemplar has some similarities to a proposal by Jones (1982), in response to Osherson and Smith (1981).

maximal typicality for C2 (the modifier) among the exemplars that are most typical for C1 (the head noun).

pet fish
Given the concept *pet*, let us assume that the most typical exemplars are *cat* and *dog*, followed by *rabbit* and *parrot*. Now assume that these instances are more typical than a *guppy*. With the concept *fish*, the instances *cod*, *salmon* and *trout* are probably more typical than a *guppy*. When we now combine the concept *pet* with the concept *fish*, there is clearly a conflict in typicality. While *pet* prefers soft and cuddly animals like cats and dogs, *fish* prefers animals that live in the sea and are caught in nets like cods and trouts. In sum (with some examples for the sake of the argument):

concepts	typicality
C1: *fish*	cod > salmon > guppy > dragonfish
C2: *pet*	dog > parrot > guppy > snake

If we use Winter's proposal or a modified MTH to solve this conflict, then the 'core exemplar' for *pet fish* is the maximally typical exemplar of a *pet* among the exemplars that are most typical for *fish*. That compromise correctly picks out a guppy. Therefore, a guppy is categorized as a *pet fish*.

red hair
The most typical hues for the concept *red* are those that are closest to focal red. Focal red is definitely a more typical hue for *red* than orange-type hues that we call 'ginger' in the context of hair. With the concept *hair*, instances of brown hair and blonde hair are probably more typical than ginger hair. Just like with *pet fish*, it is immediately clear that when we combine the two individual concepts, there is a conflict in typicality. While *red* prefers focal red hues, *hair* prefers brown or blonde hues. In sum (with some examples for the sake of the argument):

concepts	typicality
C1: *hair*	brown hair > black hair > ginger hair > blue hair
C2: *red*	focal red > dark red > orange/'ginger' > brown

The compromise interpretation of *red hair* according to the MTH and according to Winter's account is indeed 'ginger' hair, because it is the maximally typical hue of *red* among the examplars that are most typical for *hair*. Therefore, instances of ginger hair are categorized as *red hair*.

6.4 Conclusion

In this chapter we have seen that three different cases of concept composition can be described as involving conflicts of a similar kind. Winter's recent proposal, if further supported, might lead to a new theoretical avenue of describing the resolutions of these conflicts, as resulting from one uniform mechanism that generalizes the MTH. While much research is still needed in this domain, further theoretical advances will have to rely on more experimental work, exploring possible correlations between the interpretation of plural VPs, as studied in this dissertation, and the interpretation of modification constructions. If successful, there is reason to hope that such an attempt might lead to a more general theory of concept composition in language.

Chapter 7
Conclusion

This dissertation studied two types of plural sentences: reciprocal sentences and sentences with predicate conjunction. Results of the experiments that were reported show that interpretations of these constructions vary systematically between speakers and between situations. The general conclusion from these results is that content words affect the logical interpretation of complex plural sentences, unlike what traditional formal semantic studies of sentence meaning expect. The main theoretical question throughout the work has been what mechanism accounts for the interaction between logical sentence interpretation and non-logical meanings of content words. I have argued that this interaction is governed by a new principle, the Maximal Typicality Hypothesis (MTH), which is formulated using insights on concepts from cognitive psychology. The new experimental work on plural reciprocal and plural conjunctive constructions in this dissertation supports the MTH. In this chapter I look back on the main proposals and findings, and suggest directions for further research.

The MTH is described as a mechanism which works in plural reciprocal and conjunctive sentences. It uses conceptual information on content words to determine how acceptable a plural sentence will be in a given situation. The conceptual information that the MTH has recourse to pertains to speakers' perceived *typicality* of different situations for the concepts that content words refer to. On the basis of that information, the MTH singles out one situation as the *core situation* that is described by a sentence. The core situation is an optimal 'compromise': the situation that is maximal for the sentence among the situations that are most typical for the concept(s) within it. I used the MTH for explaining the acceptability of plural sentences in different situations. Firstly, plural sentences were expected to be highly acceptable in their core situation as well as in situations that properly contain the core situation (in so far as the latter are possible). Secondly, plural sentences were expected to be less acceptable in situations that are properly contained in the core situation. These predictions of the MTH were tested experimentally by first measuring what is perceived as typical for a given concept in order to predict the core situation of a sentence, and then approximating sentence acceptability in different situations.

The first type of constructions that were investigated involved plural reciprocal sentences such as *John, Bill and George are pinching each other*. The MTH selects as the core situation for such sentences the maximal situation among the situations that are most typical for the verb concept in the scope of the reciprocal expression. I used the MTH to make predictions about acceptability of a reciprocal sentence in its core situation as well as non-core situations. Experiment 1 (reported in Chapter 2) tested those predictions. Part 1 of Experiment 1 measured what is typical for verb concepts in isolation, by comparing the typicality of two different situations: situations with one patient per agent vs. situations with two simultaneous patients per agent. This part revealed that verb concepts give rise to typicality effects, just like concepts expressed by nouns are known to do. While for some verbs, a situation with one patient per agent was considered more typical than a situation with two simultaneous patients per agent, other verbs did not show differences in typicality between these two situations. From those preferences on simple situations, I extrapolated typicality preferences for more complex situations, i.e. some verbs have a preference for *S3* (in which three agents each have one patient) over *S6* (in which three agents each have two patients) while others do not have any preference. Based on those extrapolations from Part 1, I made predictions about the core situations of different reciprocal sentences with three agents. Part 2 of Experiment 1 measured the acceptability of these reciprocal sentences in two different situations (*S3* and *S6*), and revealed that acceptability varied across sentences. This variation was explained by the typicality properties that were characterized for verbs in Part 1, as supported with a correlation analysis between Part 1 and Part 2. These results support the MTH. In doing so, the results demonstrated a substantial advance over the previous attempts to test an initial version of the MTH in Kerem, Friedmann and Winter (2009).

The second type of sentences that I investigated involved plural sentences with predicate conjunction such as *The animals are big and gray* or *The men are sitting and cooking*. These sentences were investigated both in behavioral experiments (Chapter 3) and in a neurolinguistic study (Chapter 4). Firstly, Experiment 2 (reported in Chapter 3) tested the predictions of the extended Strongest Meaning Hypothesis (eSMH; Winter, 2001a). The eSMH predicts sentences with predicate conjunction to receive one of two logical interpretations (intersective or non-intersective) based on properties of the conjoined predicates. Experiment 2 revealed that sentences with predicate conjunction are accepted in a situation that is compatible with a non-intersective interpretation much more often when the predicates are incompatible (e.g. *big and small*) than when they are compatible (e.g. *big and gray*) – in line with Winter's hypothesis. However, the results also revealed differences between sentences that the eSMH would treat identically. I hypothesized that this is because the eSMH should be replaced by the MTH, which is a more general principle that takes into account subtle differences in typicality effects. For sentences with predicate conjunction, the core situation that is selected by the MTH is the maximal situation for the sentence among the situations that are most typical for the conjoined concepts. Again, I used the MTH to make predictions about acceptability of a sentence in its core situation as well as non-core situations. These predictions were tested in Experiment 3 in Chapter 3. Part 1 of Experiment 3 measured the typicality of one particular situation for different verb concept combinations, namely one where both verbs apply to one individual simultaneously. This part revealed that preferences vary across verb concept combinations. While for some verbs, simultaneous application is quite typical, for others it is not. From those preferences, I extrapolated typicality preferences for more complex situations: some verb combinations prefer split situations (in which each one of four individuals performs only one action) over joint situations (in which each one of four individuals performs two actions simultaneously), while others do not have any preference. From those extrapolations, I made predictions about the core situations of different sentences with four agents. Part 2 of Experiment 3 measured the acceptability of those sentences in a split situation and revealed that their acceptability varied. Just like with reciprocal sentences in Experiment 1, this variation was explained by the typicality differences as measured in Part 1. This was supported with a correlation analysis between Part 1 and Part 2, providing evidence for the MTH in this domain as well. This marks a substantial advance in the analysis of plurals: not only is the MTH suitable for the semantics of reciprocals, as Chapter 2 established, it can also be used to replace Winter's treatment of plural predicate conjunction.

Differences between plural sentences with predicate conjunction were also revealed in Experiment 4 in Chapter 4, a neurolinguistic study using magnetoencephalography. Experiment 4 compared neural activity within the Left Anterior Temporal Lobe (LATL) - a brain region that has been linked to the composition of complex concepts - during speakers' processing of sentences with different conjoined predicates. Results of Experiment 4 revealed that the LATL area is more active when speakers process sentences that they interpret intersectively (e.g. *The hearts are green and big*) than when they process sentences that they interpret non-intersectively (e.g. *The hearts are small and big*). In terms of conceptual combination, an intersective interpretation of the conjunction implies the integration of two features to form a single coherent entity representation, while a non-intersective interpretation of conjunction does not. Hence, the obtained differences in neural activity between these two interpretations teach us something about the contribution of the LATL in concept composition. Moreover, it supports the behavioral finding from Experiment 2 and Experiment 3 that complex plural sentences show differences in interpretation. Further research is needed to investigate whether the MTH – i.e. the systematic relation between typicality and interpretation – can also be supported using neurolinguistic measures.

Other options for further research include exploring other measures of typicality effects with verb concepts. The experiments that were reported in this dissertation each used one measurement, i.e. whether one or two patients per agent are more typical (in Experiment 1) and whether simultaneous application of two predicates is typical (in

Experiment 3). For verb concepts like *pinch*, another potential relevant typicality measurement is ‘agent cardinality’, i.e. how many agents perform an action on one patient simultaneously. With verb concept combinations like *sitting and cooking*, it could be that, for instance, the typicality of the two activities occurring in sequence also has an effect on interpretation. Furthermore, besides other lexical-conceptual factors, the dissertation also touched upon a completely different factor potentially influencing plural sentence interpretation. In Chapter 5, I presented experimental work that suggests that spatial knowledge about situations also affects the way that reciprocal sentences are interpreted. In addition, I pointed out examples of predicate conjunction that also appear to be affected by spatial information. Additional research is needed to explore these effects further.

Besides zooming in deeper on plural sentences, another worthwhile avenue is to go broader by looking at the scope of a principle like the MTH. Are there other types of constructions where we see similar effects? In Chapter 6, I have pointed out that the results obtained on complex plurals in this dissertation can be described in similar terms as results on concept composition in modification constructions such as *red hair*. In all of these constructions, there are conflicts in typicality between composed items. Further research is needed to determine whether the MTH can function as a general principle describing how language use – as ultimately driven by the functions of the human brain – resolves such conflicts.

Appendices

Appendix A: Results Chapter 2

Table A1: Pretest for Experiment 1: Preference (pref.) one-patient situations.

Verb (Dutch)	English translation	Type	Pref. One-patient situation (%)
haten	**hate**	1	.762
bewonderen	**admire**	1	.762
benijden	**envy**	1	.714
missen	**miss**	1	.667
begrijpen	**understand**	1	.571
kennen	**know**	1	.143
complimenteren	compliment	1	.619
aanbidden	adore	1	.952
houden van	love	1	.850
vereren	cherish	1	.800
genieten	enjoy	1	.714
verwijzen	refer	1	.667
denken aan	think about	1	.667
noemen	name	1	.619
horen	hear	1	.524
zien	see	1	.429
steken	**stab**	2	.952
slaan	**hit**	2	.952
knijpen	**pinch**	2	.905
strelen	**caress**	2	.905
grijpen	**grab**	2	.857
schieten	**shoot**	2	.857
schudden	shake	2	.857
verblinden	dazzle	2	.857
afvegen	wipe	2	.857
schilderen	paint	2	.857
verven	brush	2	.850
krabben	scratch	2	.850
wijzen	point	2	.810
inzepen	put soap	2	.810
afdrogen	dry	2	.800
aanraken	touch	2	.800

afspoelen	rinse	2	.800
kammen	comb	2	.762
duwen	push	2	.762
schoonmaken	clean	2	.750
wassen	wash	2	.714
kietelen	tickle	2	.714
wiegen	rock	2	.714
opmaken	apply makeup	2	.650
nat maken	wet	2	.600
muziek spelen	play music	2	.600
praten	talk	2	.571
natspetteren	spray	2	.524
versieren	decorate	2	.476
kussen	**kiss**	3	.905
bijten	**bite**	3	.857
likken	**lick**	3	.810
schoppen	**kick**	3	.810
aankleden	**dress**	3	.714
trappen	**lash out**	3	.667
vastbinden	tie	3	.700
luisteren naar	listen to	3	.550
knuffelen	hug	coll	.857
tekenen	draw	coll	.810
boetseren	sculpt	coll	.789
fotograferen	photograph	coll	.571
spreken	give a speech	coll	.550
bezoeken	visit	coll	.524
ontmoeten	meet	coll	.429

Type refers to the verb type assumed *a priori* (coll = collective). The items that appear in boldface were selected for Experiment 1.

Table A2: Experiment 1: Preference (pref.) one-patient situations and acceptance (acc.) *S2*, *S3* and *S6*.

Verb (Dutch)	English translation	Type	Pref. One-patient situation (%)	Acc. of *S2* (%)	Acc. of *S3* (%)	Acc. of *S6* (%)
benijden	envy	1	.333	.080	.750	.960
kennen	know	1	.333	.080	.360	1.00
begrijpen	understand	1	.389	.040	.440	.960
bewonderen	admire	1	.389	.040	.480	.960
missen	miss	1	.444	.125	.680	1.00
haten	hate	1	.667	.000	.440	1.00
knijpen	pinch	2	.667	.160	.880	.880
slaan	hit	2	.778	.120	.880	.840
strelen	caress	2	.833	.080	.840	.880
steken	stab	2	.882	.120	.960	.800
beschieten	shoot	2	.889	.120	.880	.840
grijpen	grab	2	.944	.160	.840	.880
kussen	kiss	3	.647	.120	.720	.640
(aan)kleden	dress	3	.889	.120	.840	.600
schoppen	kick	3	.889	.120	.840	.680
trappen	lash out	3	.944	.120	.840	.640
bijten	bite	3	1.00	.125	.840	.440
likken	lick	3	1.00	.120	.880	.520

Appendix B: Results Chapter 3

Table B1: Results Experiment 2.

Adjectives (Dutch)	English translation	Type	% "true" judgment (st.dev.)
Groot en grijs	Big and gray	Comp	.00 (.00)
Lang en zwart	Long and black	Comp	.07 (.27)
Gestreept en rood	Striped and red	Comp	.40 (.51)
Hoog en rood	High and red	Comp	.20 (.41)
Roodharig en boos	Redhaired and angry	Comp	.20 (.41)
Kort en zwartharig	Short and blackhaired	Comp	.20 (.41)
Blauwogig en zwartharig	Blue-eyed and blackhaired	Comp	.07 (.27)
Dun en bebrild	Thin and glasses-wearing	Comp	.00 (.00)
Kaal en boos	Bald and angry	Comp	.07 (.27)
Langharig en bebrild	Longhaired and glasses-wearing	Comp	.27 (.46)
Doorgetrokken en rood	Solid and red	Comp	.13 (.35)
Driehoekig en zwart	Triangular and black	Comp	.14 (.36)
Rond en groen	Round and green	Comp	.20 (.41)
	MEAN	**Compatible**	**.15 (.29)**

Adjectives (Dutch)	English translation	Type	% "true" judgment (st.dev.)
Groot en klein	Big and small	Incomp	1.00 (0.00)
Lang en kort	Long and short	Incomp	.80 (.41)
Gestreept en effen	Striped and plain	Incomp	.79 (.43)
Hoog en laag	High and low	Incomp	.71 (.47)
Roodharig en blondharig	Redhaired and blondhaired	Incomp	.79 (.43)
Kort en lang	Short and tall	Incomp	.64 (.50)
Blauwogig en bruinogig	Blue-eyed and brown-eyed	Incomp	.93 (.26)
Dun en dik	Thin and fat	Incomp	.87 (.35)
Kaal en behaard	Bald and with-hair	Incomp	.67 (.49)
Langharig en kortharig	Longhaired and shorthaired	Incomp	.79 (.43)
Doorgetrokken en gestippeld	Solid and dotted	Incomp	.71 (.47)
Driehoekig en rechthoekig	Triangular and rectangular	Incomp	.93 (.26)
Rond en vierkant	Round and square	Incomp	.86 (.35)
	MEAN	**Incompatible**	**.81(.32)**

Table B2: Results Experiment 3.

Verbs (Dutch)	English translation	Type	% “true” judgment (st.dev.)	incompatibility rating (1-6) (st.dev.)
Zitten en lezen	Sitting and reading	Comp	.65 (.49)	1.03 (.17)
Zwaaien en lachen	Waving and smiling	Comp	.31 (.48)	1.06 (.24)
Lopen en zingen	Walking and singing	Comp	.24 (.44)	1.18 (.39)
Kruipen en schreeuwen	Crawling and screaming	Comp	.81 (.40)	2..55 (1.52)
Staan en lezen	Standing and reading	Comp	.65 (.49)	1.91 (.91)
Lezen en glimlachen	Waving and smiling	Comp	.31 (.48)	1.03 (.17)
Liggen en rekken	Lying down and stretching	Comp	.65 (.49)	1.15 (.36)
Tekenen en gapen	Drawing and yawning	Comp	.75 (.45)	1.18 (.39)
Zwemmen en lachen	Swimming and smiling	Comp	.24 (.44)	1.42 (.87)
Smsen en fronsen	Texting and frowning	Comp	.56 (.51)	1.06 (.24)
Breien en zingen	Knitting and singing	Comp	.71 (.47)	1.24 (.56)
Slapen en kwijlen	Sleeping and drooling	Comp	.63 (.50)	1.42 (.66)
Zitten en staan	Sitting and standing	Incomp	.94 (.25)	5.91 (.38)
Zwaaien en klappen	Waving and clapping	Incomp	.82 (.39)	5.27 (1.44)
Lopen en zwemmen	Walking and swimming	Incomp	.87 (.34)	5.94 (.24)
Kruipen en springen	Crawling and jumping	Incomp	.71 (.47)	5.82 (.73)
Staan en hurken	Standing and squatting	Incomp	.94 (.25)	5.85 (.51)
Lezen en slapen	Reading and sleeping	Incomp	.76 (.44)	5.88 (.41)
Liggen en rennen	Lying down and running	Incomp	.87 (.34)	5.91 (.29)
Tekenen en typen	Drawing and typing	Incomp	.82 (.39)	5.09 (.98)
Zwemmen en kruipen	Swimming and crawling	Incomp	1 (.00)	5.91 (.29)
Smsen en breien	Texting and knitting	Incomp	.71 (.47)	5.03 (1.47)
Breien en klappen	Knitting and clapping	Incomp	.87 (.34)	5.70 (.64)
Slapen en bellen	Sleeping and telephoning	Incomp	.76 (.44)	5.61 (.97)
Zitten en koken	Sitting and cooking	Atyp	.88 (.33)	3.12 (1.05)
Zwaaien en tekenen	Waving and drawing	Atyp	.75 (.45)	3.79 (1.34)
Lopen en schrijven	Walking and writing	Atyp	.81 (.40)	3.39 (1.09)
Kruipen en lezen	Crawling and reading	Atyp	.71 (.47)	4.55 (.94)
Staan en dutten	Standing and falling asleep	Atyp	.47 (.51)	3.70 (1.38)
Lezen en tekenen	Reading and drawing	Atyp	.94 (.25)	4.18 (1.31)

Liggen en drinken	Lying down and drinking	Atyp	.81 (.40)	3.42 (1.39)
Tekenen en lopen	Drawing and walking	Atyp	.76 (.44)	3.76 (1.15)
Zwemmen en lezen	Swimming and reading	Atyp	.76 (.44)	4.97 (1.05)
Smsen en zwaaien	Texting and waving	Atyp	.81 (.40)	2.39 (1.27)
Breien en lopen	Knitting and walking	Atyp	.87 (.34)	3.70 (.88)
Slapen en staan	Sleeping and standing	Atyp	.76 (.44)	4.91 (.63)

Appendix C: Experiment Chapter 5

Testing the effects of spatial knowledge on the interpretation of reciprocal sentences

This experiment measured the interpretation of Dutch plural reciprocal sentences with symmetric verbs in order to test the effects of spatial knowledge on sentence interpretation. It compared the acceptability rating of reciprocal sentences in two situations that differ only in terms of their spatial configuration.

Participants

Twenty-nine students from Utrecht University (27 female, age M = 21) participated for monetary compensation. All participants were native speakers of Dutch without dyslexia.

Materials

Eleven different Dutch reciprocal sentences were tested on their acceptability in two situations. I used two different versions of the experiment, each containing 11 unique test items plus 12 filler items that were the same across the two versions. Each test item contained a plural reciprocal sentence containing a symmetric verb of the form *A, B, and C are P-ed to each other* (where *A*, *B*, *C* and *D* visibly refered to three materials in the image and *P* is a symmetric verb), plus an image depicting the three materials in one of two situations: a line situation or a circle situation. In the line situation, the materials are situated in a row next to each other. In the circle situation, they are situated in a circular (or - sometimes more accurately - triangular) situation. In both types of situations, there are two symmetric relations among the three materials, thus creating a line or an ‘incomplete circle’. The independent variable that was manipulated is whether three symmetric relations are possible in the situation (i.e. circle) or not (i.e. line). Images were kept as equal as possible in all other respects. Examples of both types of test images for one sentence are given in Figure 1.

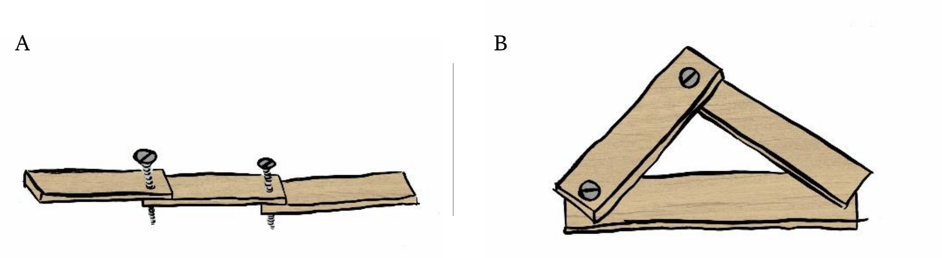

Figure 1: Test images for *The wooden planks A, B and C are screwed to each other*: line configuration (A) and circle configuration (B).

In each questionnaire, half[41] of the test items contained sentences accompanied by a line situation (as in Figure 1A) and the other half contained sentences accompanied by a circle situation (as in Figure 1B). The same eleven test sentences were used in both versions of the questionnaire. To ensure that participants never saw the same sentence twice, one of the situations occurred in version 1 and the other situation occurred in version 2. This allowed me to compare acceptability rating of the same sentence in minimally different situations. Table C1 provides an overview of the eleven symmetric verbs plus the type of materials that they were combined with within test sentences. The two rightmost columns show which situation occurred in which version.

Table C1: Overview of test items, translated from Dutch.

Symmetric verb	Materials	Version 1	Version 2
Staple	Pieces of paper	Line	Circle
Join	Points	Line	Circle
Screw	Wooden planks	Circle	Line
Sew	Pieces of fabric	Circle	Line
Pin	Small pieces of papier	Line	Circle
Hook	Ovals	Circle	Line
Glue	Pieces of cardboard	Circle	Line
Knot	Wires	Line	Circle
Knit	Pieces of knit work	Circle	Line
Nail	Wooden planks	Line	Circle
Tie	Ropes	Circle	Line

Filler items also contained reciprocal sentences with symmetric verbs, but either a) sentences about two instead of three materials (*A and B are P-ed to each other*) or b) sentences with three materials with an accompanying image where there is only one symmetric relation among three materials. The spatial configurations of filler item images were also varied. Half of the filler items were expected to be judged true, and half of them were expected to be judged false. Both versions of the questionnaire contained the same filler items.

Procedure

Each participant completed one of two questionnaires on paper, in a sound-proof booth. They were instructed to rate as soon as possible how well the sentence describes the

[41] Since the number of test items was uneven in each questionnaire, the division was five vs. six.

image on a scale of 1 to 5 (from very badly to very well), by circling a number with a pen.[42]

Coding and analysis

Responses were coded '1' through '5', corresponding to the participants' judgment. I computed the acceptability rating for each sentence in each of the two types of images (line situation and circle situation) for each participant. I then performed a repeated measures ANOVA with Situation as the within-subjects factor (with 2 levels: line and circle) and Version as between-subjects factor (with 2 levels: version 1 and version 2).

Results

Firstly, the ANOVA revealed no effect of Version (F (1,27) = .002, p = .97) nor an interaction effect of Situation*Version (F (1,27) = .21, p = .65). This indicates that the two versions of the questionnaire do not differ from each other, and thus measure the same thing (see Figure 2).

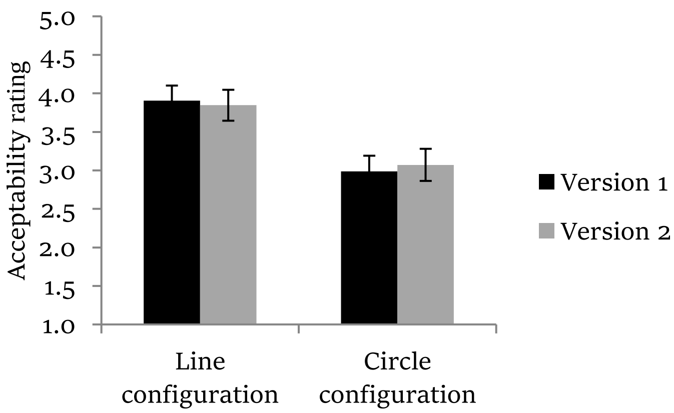

Figure 2: Results

Secondly, the ANOVA did reveal a significant main effect of Situation (F (1,27) = 29.61, p < .001). Sentences received a higher acceptability rating in the line situation than in the circle situation (see Table C2).

[42] Unlike other experiments in this dissertation, I decided to use a more sensitive measure – i.e. a scale - than a truth-value judgement task. This is because the experimental work in Poortman (2011) revealed that the differences in acceptability of reciprocal sentences with symmetric verbs in different spatial configurations are very subtle.

Table C2: Mean acceptability rating.

Situation		Mean rating (st. dev.)
Line	Version 1	3.91 (.68)
	Version 2	3.85 (.82)
	Total mean	3.88 (.74)
Circle	Version 1	2.99 (.93)
	Version 2	3.07 (.57)
	Total mean	3.03 (.77)

Bibliography

Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., & Ochiai, M. (2001). Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. *IEEE Trans. Appl. Supercond 11*, 669-672.

Baron, S. G., & Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. *Neuroimage 55*, 1847-1852.

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

Beck, S. (2001). Reciprocals are definites. *Natural Language Semantics 9*, 69-138.

Beeman, M. J., & Chiarello, C. (1998). Complementary right- and left-hemisphere language comprehension. *Current Directions in Psychological Science 7*(1), 2-8.

Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *Journal of Neuroscience 31*(8), 2801-2814.

Bemis, D. K., & Pylkkänen, L. (2013a). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex 23*(8), 1859-1873.

Bemis, D. K., & Pylkkänen, L. (2013b). Flexible composition: MEG evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands. *PLoS ONE 8*(9), e73949.

Bennett, M. (1974). Some extensions of a Montague fragment of English. Doctoral dissertation, MIT, Cambridge MA.

Bergmann, M. (1982). Cross-categorial semantics for conjoined common nouns. *Linguistics and Philosophy 5*(3), 399-401.

Blanco-Elorrieta, E., & Pylkkänen, L. (2016). Composition of complex numbers: Delineating the computational role of the left anterior temporal lobe. *NeuroImage 124*, 194-203.

Brennan, J., & Pylkkänen, L. (2012). The time-course and spatial distribution of brain activity associated with sentence processing. *Neuroimage 60*, 1139-1148.

Brisson, C. (1998). Distributivity, Maximality, and Floating Quantifiers. Doctoral dissertation, Rutgers, New Brunswick, NJ.

Burnett, H. (2011). Vagueness and Plural Predication. In D. Lassiter (Ed.) *Proceedings of the 2011 ESSLLI Student Session.* Ljubljana, Slovenia.

Caramazza, A., Gordon, J., Zurif, E.B., & DeLuca, D. (1976). Right-hemispheric damage and verbal problem solving behavior. *Brain and Language 3*, 41-46.

Casteren, M. van, & Davis, M.H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods 38*(4), 584-589.

Champollion, L. (2015). Ten men and women got married today: noun coordination and the intersective theory of conjunction. *Journal of Semantics 0*, 1–62.

Chomsky, N. (1973). Conditions on transformations. In S. Anderson, & P. Kiparsky (Eds.), *A Festschrift for Morris Halle*, 232-286. Academic Press, New York.

Chomsky, N. (1986). Knowledge of language: its nature, origin and use. Praeger, New York.

Dalrymple, M., Kanazawa, M., Kim, Y., Mchombo, S., & Peters, S. (1998). Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy 21*, 159-210.

Davidson. D. (1967). The logical form of action sentences. In N. Rescher (Ed.), *The logic of decision and action,* 81-94. University of Pittsburgh Press, Pittsburgh. Reprinted in Davidson (1980), 105-122.

Del Prato, P., & Pylkkänen, L. (2014). MEG evidence for conceptual combination but not numeral quantification in the left anterior temporal lobe during language production. *Frontiers in Psychology 5* (524).

Done, D. J., & Gale, T. M. (1997). Attribute verification in dementia of Alzheimer type: Evidence for the preservation of distributed concept knowledge. *Cognitive Neuropsychology 14*(4), 547-571.

Dougherty, R.C. (1974). The syntax and semantics of *each other* constructions. *Foundations of Language 12*, 1-47.

Dowty, D. (1988). Type raising, functional composition, and non-constituent conjunction. In R. Oehrle et al. (Eds.), *Categorial Grammars and Natural Language Structures*. Reidel, Dordrecht.

Ferstl, E. C., Neumann, J., Bogler, C., & Von Cramon, D. Y. (2008). The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Human brain mapping 29*(5), 581-593.

Fiengo, R., & Lasnik, H. (1973). The logical structure of reciprocal sentences in English. *Foundations of Language 9*(4), 447-468.

Friederici, A. D., Meyer, M., & Von Cramon, D. Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language 74*(2), 289-300.

Gardner, H., Silverman, J., Wapner, W., & Zurif, E. (1978). The appreciation of antonymic contrasts in aphasia. *Brain and Language 6*, 301-317.

Gazdar, G. (1980). A cross-categorial semantics for coordination. *Linguistics and Philosophy 3*, 407-409.

Genovese, C.R., Lazar, N.A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage 15* (4), 870–878.

Grice, H.P. (1975). Logic and conversation. In P. Cole, & J.L. Morgan (Eds.), *Syntax and semantics 3: Speech acts*. Academic Press, New York.

Hampton, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition 14*(1), 12–32.

Hampton, J.A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science 31*(3), 355-384.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar* (Vol. 13). Blackwell, Oxford.

Heim, I., Lasnik, H., & May, R. (1991). Reciprocity and plurality. *Linguistic Inquiry 22*, 63-101.

Heycock, C. & Zamparelli, R. (2005). Friends and colleagues: plurality, coordination, and the structure of DP. *Natural Language Semantics 13*, 201-270.

Hodges, J. R., Graham, N., & Patterson, K. (1995). Charting the progression in semantic dementia: Implications for the organization of semantic memory. *Semantic Knowledge and Semantic Representations 3*(3-4), 463-495.

Hoeksema, J. (1983). Plurality and conjunction. In A. ter Meulen (Ed.), *Studies in Modeltheoretic Semantics*. Foris, Dordrecht.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience 18*(4), 665-679.

Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human Brain Mapping 26*(2), 128-138.

Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport 12*(8), 1749-1752.

Janssen, T. M. V. (1997). Compositionality. In J. Van Benthem, & A. ter Meulen (Eds.), *Handbook of Logic and Language*, Elsevier, Amsterdam.

Jobard, G., Vigneau, M., Mayozer, B., & Tzourio-Mayozer, N. (2007). Impact of modality and linguistic complexity during reading and listening tasks. *Neuroimage 34*, 784-800.

Jones, G. V. (1982). Stacks not fuzzy sets: An ordinal basis for prototype theory of concepts. *Cognition 12*, 281-290.

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

Kański, Z. (1987). Logical symmetry and natural language reciprocals. *Proceedings of the 1987 Debrecen Symposium on Language and Logic*, 49-68. Akadémiai Kiadó, Budapest.

Keenan, E., & Faltz, L. (1985). *Boolean Semantics for Natural Language*. D. Reidel, Dordrecht.

Kerem, N., Friedmann, N., & Winter, Y. (2009). Typicality effects and the logic of reciprocity. In E. Cormany, S. Ito, & D. Lutz (Eds.), *Proceedings of Semantics and Linguistic Theory, SALT19*, 257-274. eLanguage.

King, T.H., & Dalrymple, M. (2004). Determiner agreement and noun conjunction. *Journal of Linguistics 40*(1), 69-104.

Krifka, M. (1990). Boolean and non-boolean ‘and’. In L. Kálmán, & L. Pólós (Eds.), *Papers from the Second Symposium of Logic and Language.* Akadémiai Kiadó, Budapest.

Križ, M. (2015). Homogeneity, Non-Maximality, and *all. Journal of Semantics.*

Lambon Ralph, M. A., Cipolotti, L., Manes, F., & Patterson, K. (2010). Taking both sides: Do unilateral anterior temporal lobe lesions disrupt semantic memory? *Brain 133*(11), 3243-3255.

Lambon Ralph, M.A., Pobric, G., & Jefferies, E. (2009). Conceptual knowledge is underpinned by the temporal pole bilaterally: convergent evidence from rTMS. *Cereb. Cortex 19* (4), 832.

Lancaster, J., Rainey, L., Summerlin, J., Freitas, C., Fox, P., Evans, A., et al. (1997). Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Human Brain Mapping 5*(4), 238.

Lancaster, J., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping 10*(3), 120-131.

Langendoen, D. T. (1978). The logic of reciprocity. *Linguistic Inquiry 9*(2), 177-197.

Lasersohn, P. (1988). A Semantics for Groups and Events. Doctoral dissertation, Ohio State University, Columbus.

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

Lebeaux, D. (1983). A distributional difference between reciprocals and reflexives. *Linguistc Inquiry 14*, 723-730.

Lee, C. (to appear). Typicality knowledge and the interpretation of adjectives. To appear in J. Hampton, & Y. Winter (Eds.), *Compositionality and Concepts in Linguistics and Psychology*. Springer.

Leffel, T., Lauter, M., Westerlund, M., & Pylkkänen, L. (2014). Restrictive vs. non-restrictive composition: a magnetoencephalography study. *Language, cognition and neuroscience 29*(10), 1191-1204.

Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bauerle et al. (Eds.), *Meaning, Use and Interpretation of Language*. De Gruyter, Berlin.

Link, G. (1984). Hydras: On the logic of relative constructions with multiple heads. In F. Landman, & F. Veltman (Eds.), *Varieties of Formal Semantics*. Foris, Dordrecht.

Link, G. (1987). Generalized quantifiers and plurals. In. P. Gärdenfors (Ed.), *Generalized Quantifiers*. D. Reidel, Dordrecht.

Margolis, E., & Laurence, S. (1999). Concepts and cognitive science. In E. Margolis, & S. Laurence (Eds.), *Concepts: Core readings*, 3-82. MIT Press, Cambridge, MA.

Mari, A. (2013). *Each Other*, asymmetry and reasonable futures. *Journal of Semantics 31*(2), 209-261.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods 146* (1), 177–190.

Massey, G.J. (1976). Tom, Dick, and Harry, and all the king's men. *American Philosophical Quarterly 13*, 89-107.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods 44*(2), 314-324.

Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., et al. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience 5*(4), 467-479.

Montague, R. (1970). Universal grammar. *Theoria* 36, 373-98.

Osherson, D., & Smith, E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition 9*, 35–58.

Pallier, C., Devauchelle, A. D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences 108*(6), 2522-2527.

Parsons, T. (1990). *Events in the Semantics of English* (Vol. 5). MIT Press, Cambridge MA.

Partee, B., & Rooth, M. (1983). Generalized conjunction and type ambiguity. In R. Bauerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, Use and Interpretation of Language*. De Gruyter, Berlin.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience 8*(12), 976-987.

Philip, W. (2000). Adult and Child Understanding of Simple Reciprocal Sentences. *Language 76*(1), 1-27.

Poortman, E.B. (2011). Reciprocals and Geometry-sensitivity: An experimental study. MPhil Thesis, Utrecht University.

Poortman, E.B. (2014). Between Intersective and 'Split' Interpretations of Predicate Conjunction: The Role of Typicality. In J. Degen, M. Franke, & N. Goodman (Eds.), *Proceedings of the Formal & Experimental Pragmatics Workshop*, 36-42, Tübingen.

Poortman, E.B. (to appear). Concept typicality and the interpretation of plural predicate conjunction. To appear in J. Hampton, & Y. Winter (Eds.), *Compositionality and Concepts in Linguistics and Psychology*. Springer.

Poortman, E.B., & Pylkkänen, L. (2016). Adjective conjunction as a window into the LATL's contribution to conceptual combination. *Brain and Language 160*, 50-60.

Poortman, E.B., Struiksma, M.E., Kerem, N., Friedmann, N., & Winter, Y. (under revision). Reciprocal expressions and the Maximal Typicality Hypothesis.

Rice, G. E., Ralph, M. A. L., & Hoffman, P. (2015). The Roles of Left Versus Right Anterior Temporal Lobes in Conceptual Knowledge: An ALE Meta-analysis of 97 Functional Neuroimaging Studies. *Cerebral Cortex*, bhv024.

Rogalsky, C., & Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex 19*(4), 786-796.

Rogers, T. T., Hocking, J., Noppeney, U., Mechelli, A., Gorno-Tempini, M. L., Patterson, K., et al. (2006). Anterior temporal cortex and semantic memory: Reconciling findings from neuropsychology and functional imaging. *Cognitive, Affective, & Behavioral Neuroscience 6*(3), 201-213.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review 111*(1), 205.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*, 111-144. Academic Press, New York.

Rosch, E., & C. B. Mervis (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology 7*, 573-605.

Sabato, S. (2006). The semantics of reciprocal expressions in natural language. Msc. Thesis.

Sabato, S., & Winter, Y. (2012). Relational domains and the interpretation of reciprocals. *Linguistics and Philosophy 25*(3), 191-241.

Sassoon, G.W. (to appear). Dimension accessibility as a predictor of morphological gradability. To appear in J. Hampton, & Y. Winter (Eds.), *Compositionality and Concepts in Linguistics and Psychology*. Springer.

Schwarz, F. (2013). Maximality and Definite Plurals – Experimental evidence. In E. Chemla, V. Homer, & G. Winterstein (Eds.), *Proceedings of Sinn und Bedeutung 17*, 509-526.

Schwarzschild, R. (1996). Pluralities. Kluwer, Dordrecht.

Smith, E.E., Shoben, E.J., & Rips, L.J. (1974). Structure and Process in Semantic Memory: A Featural Model for Semantic Decisions. *Psychological Review 81* (3), 214-241.

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

St George, M., Kutas, M., Martinez, A., & Sereno, M.I. (1999). Semantic integration in reading: engagement of the right hemisphere during discourse processing. *Brain 122*, 1317–1325.

Storms, G., de Boeck, P., van Mechelen, I., & Ruts, W. (1998). Not guppies, nor goldfish, but tumble dry-ers, Noriega, Jesse Jackson, panties, car crashes, bird books, and Stevie Wonder. *Memory & cognition 26*(1), 143–145.

Stowe, L. A., Broere, C. A., Paans, A. M., Wijers, A. A., Mulder, G., Vaalburg, W., & Zwarts, F. (1998). Localizing components of a complex task: sentence processing and working memory. *Neuroreport 9*(13), 2995-2999.

Tieu, L., Romoli, J., Poortman, E.B., Winter, Y., & Crain, S. (under revision). Children's comprehension of plural predicate conjunction.

Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience 14*(4), 550-560.

Visser, M., Jefferies, E., & Ralph, M. L. (2010). Semantic processing in the anterior temporal lobes:a meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, *22*(6), 1083-1094.

Westerlund, M., & Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia 57*, 59-70.

Westerlund, M., Kastner, I., Al Kaabi, M., & Pylkkänen, L. (2015). The LATL as locus of composition: MEG evidence from English and Arabic. *Brain and Language 141*, 124–134.

Westerstahl, D. (1984). Determiners and context sets. In J. van Benthem & A. ter Meulen (Eds.), *Generalized Quantifiers in Natural Language*, 45-71. Foris, Dordrecht.

Winter, Y. (1998). Flexible Boolean semantics: coordination, plurality and scope in natural language. Doctoral dissertation, Utrecht University, Netherlands.

Winter, Y. (2001a). Plural predication and the strongest meaning hypothesis. *Journal of Semantics 18*, 333-365.

Winter, Y. (2001b). *Flexibility principles in Boolean semantics: the interpretation of coordination, plurality, and scope in natural language.* MIT Press, Cambridge MA.

Winter, Y. (to appear). Critical typicality: truth judgements and compositionality with plurals and other gradable concepts. To appear in J. Hampton, & Y. Winter (Eds.), *Compositionality and Concepts in Linguistics and Psychology*. Springer.

Zhang, L., & Pylkkänen, L. (2015). The interplay of composition and concept specificity in the left anterior temporal lobe: An MEG study. *NeuroImage* 111, 228-240.

Samenvatting in het Nederlands

Dit proefschrift gaat over de interpretatie van twee typen meervoudzinnen: zinnen met een wederkerig voornaamwoord zoals in (1) en (2), en zinnen met nevenschikking van twee predicaten zoals (3) en (4).

(1) John, Bill en George knijpen elkaar
(2) John, Bill en George kennen elkaar

(3) De mannen zitten en koken
(4) De mannen zitten en lezen

Experimentele resultaten in dit proefschrift tonen aan dat de interpretaties van dergelijke constructies systematisch variëren, ook al verschillen ze enkel wat betreft inhoudswoorden (zoals *knijpen* vs. *kennen*, en *koken* vs. *lezen*). Zo wordt zin (3) bijvoorbeeld vaak als 'waar' beoordeeld in een situatie met vier mannen waarin twee van de mannen zitten en de overige twee mannen koken, terwijl zin (4) veel minder vaak als 'waar' wordt beoordeeld in de logisch equivalente situatie waarin twee mannen zitten en twee mannen lezen. De algemene conclusie die ik uit deze resultaten trek, is dat inhoudswoorden een effect hebben op de logische interpretatie van meervoudzinnen. Dat gaat tegen de verwachtingen van traditionele formeel semantische studies van zinsbetekenis in. De belangrijkste theoretische doelstelling van het proefschrift was het poneren van een mechanisme dat de interactie verklaart tussen de logische interpretatie van een zin aan de ene kant, en de niet-logische betekenis van inhoudswoorden aan de andere kant. Ik beweer dat deze interactie verloopt volgens een nieuw principe, de *Maximal Typicality Hypothesis* (MTH), dat gebruik maakt van inzichten over concepten uit de cognitieve psychologie. Mijn nieuwe experimentele resultaten over meervoudzinnen met wederkerige voornaamwoorden en meervoudzinnen met nevenschikking van predicaten leveren steun voor de MTH.

Ik beschrijf de MTH als een mechanisme dat werkzaam is in meervoudzinnen met wederkerige voornaamwoorden en met nevenschikking van predicaten. De MTH gebruikt conceptuele informatie over inhoudswoorden in een meervoudzin om te bepalen hoe acceptabel die zin zal zijn in een bepaalde situatie. De specifieke conceptuele informatie waar de MTH een beroep op doet, is zogenaamde *typicaliteit* (typicality). Cognitieve psychologen hebben ontdekt dat de mentale concepten waar woorden naar verwijzen niet alleen informatie bevatten over categorisatie binnen een concept (is een *pinguïn* wel of geen lid van het concept *vogel*?), maar ook over hoe typisch verschillende leden zijn binnen een concept (hoe typisch is een *pinguïn* als lid van het concept *vogel*?). Ter illustratie, zowel een *mus* als een *pinguïn* worden door sprekers gecategoriseerd als vogel, maar daarnaast wordt een *mus* consistent ervaren als een typischere vogel dan een *pinguïn*. In dit proefschrift toon ik aan dat typicaliteitseffecten binnen concepten ook bestaan voor werkwoorden. Werkwoordconcepten categoriseren geen entiteiten (zoals *pinguïn*, *mus*), maar situaties (zoals *Jan zit*, *Jan kent Bill en George*). De MTH gebruikt typicaliteitseffecten voor werkwoorden: de door sprekers ervaren mate van typicaliteit van verschillende situaties voor gegeven werkwoordconcepten. Voor de concepten *zitten* en *koken* in zin (3) gaat dat er bijvoorbeeld om hoe typisch sprekers een situatie vinden waarin een man tegelijkertijd zit en kookt. Op basis van dat soort conceptuele informatie selecteert de MTH één *kernsituatie* die door een meervoudzin beschreven wordt. Die kernsituatie is een optimaal compromis: de situatie die maximaal passend is voor de zin, gegeven de situaties die het meest typisch zijn voor de concepten in die zin. Aan de hand van de MTH verklaar ik hoe vaak meervoudzinnen als 'waar' worden beoordeeld in verschillende situaties. Met andere woorden, de MTH voorspelt hun acceptatiepatroon. Ten eerste verwachtte ik dat meervoudzinnen zeer acceptabel zouden zijn in zowel hun kernsituatie als in situaties die de kernsituatie omvatten (voor zover die laatste situaties mogelijk zijn). Ten tweede verwachtte ik dat meervoudzinnen minder acceptabel zouden zijn in situaties die bevat worden binnen de kernsituatie. Deze verwachtingen heb ik experimenteel getest voor de twee typen meervoudzinnen. Eerst heb ik gemeten wat sprekers als typisch ervaren voor een gegeven concept om vervolgens te kunnen voorspellen wat de kernsituatie van een zin met dat concept erin is, en ten slotte heb ik gemeten hoe acceptabel een zin is in verschillende situaties.

Het eerste type zinnen dat ik onderzocht heb, is meervoudzinnen met een wederkerig voornaamwoord zoals (1) en (2). Voor dit soort zinnen selecteert de MTH als kernsituatie de beste passende (maximale) situatie voor de zin onder de situaties die het meest typisch worden bevonden voor het werkwoordconcept in de zin. Ik gebruikte de MTH om vervolgens voorspellingen te doen over de acceptatie van een zin met een wederkerig voornaamwoord in zijn kernsituatie, en in overige situaties. Experiment 1 in hoofdstuk 2 testte die voorspellingen. Deel 1 van experiment 1 mat wat sprekers typisch vinden voor een gegeven werkwoordconcept zoals *knijpen*, door de typicaliteit van twee verschillende situaties te vergelijken: situaties met één *patiënt* ('lijdend voorwerp') per *agent* (handelend persoon) vs. situaties met twee patiënten per agent tegelijk. Deel 1 liet zien dat werkwoordconcepten verschillen in hun voorkeuren. Dat wil zeggen dat ze *typicaliteitseffecten* vertoonden, net zoals is aangetoond voor concepten die worden uitgedrukt door zelfstandige naamwoorden. Voor sommige werkwoordconcepten vonden sprekers het typischer wanneer er slecht één patiënt per agent was (bijvoorbeeld *knijpen*), terwijl ze bij andere werkwoorden geen voorkeur hadden tussen de twee geteste situaties (bijvoorbeeld *kennen*). Aan de hand van deze voorkeuren wat betreft simpele situaties, extrapoleerde ik voorkeuren bij complexere situaties: voor sommige werkwoorden prefereren sprekers situaties waarin drie agenten elk één patiënt hebben (*S3*) boven situaties waarin drie agenten elk twee patiënten hebben (*S6*), terwijl ze voor andere werkwoorden geen voorkeur hebben. Volgens die conclusies uit deel 1, voorspelde ik aan de hand van de MTH welke situaties de kernsituaties zijn voor zinnen met een wederkerig voornaamwoord met drie agenten. Deel 2 van experiment 1 mat hoe acceptabel deze zinnen zijn in de twee situaties *S3* en *S6*, en toonde aan dat de mate van acceptatie varieerde tussen zinnen. Die variatie kon worden verklaard door de typicaliteitsmetingen uit deel 1. Het verband tussen typicaliteit uit deel 1 en acceptatie van zinnen in deel 2 ondersteunde ik met een correlatieanalyse tussen de twee delen. Deze resultaten steunen de MTH.

Daarna onderzocht ik een tweede type zinnen: meervoudzinnen met nevenschikking van predicaten zoals (3) en (4). Dit type zinnen onderzocht ik in zowel gedragsexperimenten (hoofdstuk 3) als in een neurolinguïstisch experiment (hoofdstuk 4). Experiment 2 in hoofdstuk 3 testte de voorspellingen van een eerdere hypothese: de *extended Strongest Meaning Hypothesis* (eSMH; Winter 2001a). De eSMH is een generalisatie van de *Strongest Meaning Hypothesis* (SMH) van Dalrymple et al. (1998), die ik in detail bespreek in hoofdstuk 2. De eSMH voorspelt dat een zin met predicaatsnevenschikking één van twee verschillende logische interpretaties zal hebben (*intersectief* of *non-intersectief*), afhankelijk van de twee predicaten in de zin. Experiment 2 toonde aan dat zinnen met adjectiefnevenschikking veel vaker worden geaccepteerd in een situatie die compatibel is met een *non-intersectieve* interpretatie wanneer de predikaten tegenstrijdig zijn (bijvoorbeeld *groot en klein*) dan wanneer ze verenigbaar zijn (bijvoorbeeld *groot en grijs*) – volgens de verwachtingen van Winter's hypothese. Desondanks lieten de resultaten ook zien dat er verschillen bestaan tussen zinnen die volgens de eSMH identiek zouden moeten zijn. Ik hypothetiseerde dat dit voorspeld zou worden als de eSMH vervangen zou worden door de MTH. De MTH is een algemener principe dat subtiele verschillen in typicaliteit meeneemt. Voor zinnen met predicaatsnevenschikking selecteert de MTH een kernsituatie: de maximale situatie voor de zin onder de situaties die het meest typisch worden bevonden voor de gekoppelde concepten in de zin. Ook hier weer gebruikte ik de MTH om voorspellingen te doen over de acceptatie van een zin in zijn kernsituatie en overige situaties. Deze voorspellingen werden getest met zinnen met nevenschikking van werkwoorden in experiment 3 in hoofdstuk 3. Deel 1 van experiment 3 mat de typicaliteit van een specifieke situatie voor verschillende combinaties van werkwoordconcepten, namelijk een waar beide werkwoorden tegelijkertijd van toepassing zijn op één individu. Dit deel liet zien dat voorkeuren varieerden tussen verschillende combinaties werkwoorden. Voor sommige combinaties vonden sprekers het heel typisch wanneer beide werkwoorden tegelijkertijd van toepassing waren (bijvoorbeeld *zitten en lezen*), terwijl dat bij andere combinaties juist niet als typisch werd ervaren (bijvoorbeeld *zitten en koken*). Aan de hand van deze voorkeuren bij simpele situaties extrapoleerde ik voorkeuren bij complexere situaties: voor sommige werkwoorden prefereren sprekers zogenaamde *gespleten* situaties (waarin elk van vier agenten slechts één actie uitvoert) boven *gezamenlijke* situaties (waarin elk van vier agenten tegelijkertijd twee acties uitvoert), terwijl ze bij andere werkwoorden geen voorkeur hebben. Gezien de conclusies uit deel 1 voorspelde ik aan de hand van de MTH welke situaties de kernsituaties zijn voor zinnen met predicaatsnevenschikking met vier agenten. Deel 2 van experiment 3 mat hoe acceptabel deze zinnen zijn in een *gespleten* situatie, en liet zien dat de mate van acceptatie varieerde tussen zinnen. Net als bij zinnen met wederkerige voornaamwoorden in experiment 1, kon die variatie worden verklaard door de typicaliteitsmetingen uit deel 1. Het verband tussen typicaliteit uit deel 1 en acceptatie van zinnen in deel 2 werd wederom ondersteund met een correlatieanalyse tussen de twee delen. Deze resultaten bewijzen dat de MTH ook in dit domein de juiste voorspellingen doet. Deze bevindingen markeren een substantiële vooruitgang in de analyse van meervoudzinnen: niet alleen is de MTH bruikbaar voor de semantiek van wederkerige voornaamwoorden, zoals bewezen in hoofdstuk 2, ook kan het gebruikt worden om Winter's hypothese over predicaatsnevenschikking te vervangen.

Experiment 4 in hoofdstuk 4 toonde ook verschillen aan tussen meervoudzinnen met predicaatsnevenschikking, ditmaal in een neurolinguïstische studie die gebruik maakte van magnetoencephalografie. Experiment 4 vergeleek neurale activiteit tijdens het verwerken van zinnen met verschilllende predicaten, binnen de linker anteriore temporale kwab ('left anterior temporal lobe', LATL) – een hersengebied waarvan beweerd is dat het betrokken is bij de compositie van complexe concepten. Resultaten van experiment 4 lieten zien dat de LATL actiever is wanneer sprekers zinnen verwerken waaraan ze een *intersectieve* interpretatie toekennen (bijvoorbeeld *The hearts are green and big*) dan wanneer ze zinnen verwerken waaraan ze een *non-intersectieve* interpretatie toekennen (bijvoorbeeld *The hearts are small and big*). Vanuit de compositie van concepten bekeken, impliceert een *intersectieve* interpretatie van een nevenchikking de integratie van twee *eigenschappen* tot één coherente entiteitsrepresentatie, terwijl een *non-intersectieve* interpretatie dat niet impliceert. De gevonden verschillen in neurale activiteit tussen de twee interpretaties leren ons dus iets over de bijdrage van de LATL aan het proces van conceptcompositie. Daarnaast ondersteunen ze de bevinding uit de gedragsexperimenten (experiment 2 en experiment 3) dat syntactisch vergelijkbare meervoudzinnen kunnen leiden tot verschillende interpretaties. Verder onderzoek is nodig om te onderzoeken of de werking van de MTH – dat wil zeggen de systematische relatie tussen typicaliteit en interpretatie – ook kan worden aangetoond met neurolinguïstische onderzoeksmethoden.

Een andere optie voor verder onderzoek werd besproken in hoofdstuk 5. Naast lexicaal-conceptuele factoren, ging het proefschrift in hoofdstuk 5 kort in op een heel andere factor die potentieel invloed uitoefent op zinsinterpretatie. In dat hoofdstuk presenteerde ik experimenteel werk dat suggereert dat ruimtelijke kennis over situaties invloed heeft op de interpretatie van zinnen met wederkerige voornaamwoorden. Daarnaast heb ik voorbeelden gegeven van zinnen met predicaatsnevenschikking waarbij ruimtelijke informatie ook een rol lijkt te spelen. Meer onderzoek is nodig om deze effecten verder te bestuderen.

Behalve verder inzoomen op meervoudzinnen, opent dit proefschrift ook de weg naar vervolgonderzoek dat juist breder kijkt naar het bereik van een principe als de MTH. Zijn er andere constructies waar we een zelfde soort effect zien? In hoofdstuk 6 heb ik laten zien dat de resultaten met betrekking tot meervoudzinnen uit dit proefschrift op eenzelfde manier beschreven kunnen worden als resultaten over conceptcompositie in modificatieconstructies zoals *rood haar*. In al deze constructies zijn er conflicten in typicaliteit tussen concepten die gecombineerd worden. Vervolgonderzoek moet uitwijzen of de MTH kan functioneren als een algemeen principle dat beschrijft hoe taalgebruik – wat uiteindelijk gedreven wordt door de functies van het menselijk brein – dit soort conflicten oplost.

Curriculum Vitae

Eva Poortman was born on October 24, 1986, in Maastricht in the Netherlands. After obtaining her VWO diploma in 2006, she studied at Utrecht University from 2006 until 2011. In 2009, she graduated with a BA in Communication and Information Sciences (cum laude), and in 2011 she obtained her MA degree in Linguistics (cum laude).

This dissertation is the result of work that Eva carried out between 2011 and 2016 as a PhD researcher at Utrecht University within the NWO-funded project 'Between Logic and Common Sense: the Formal Semantics of Words'. During that period, she worked as a visiting researcher at New York University (NYU) in the United States from January until June 2014.

Eva Poortman currently works as a lecturer in Research and Communication at de Haagse Hogeschool in The Hague, the Netherlands.