An experimental investigation of presupposition processing

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Cosima Isadora Schneider
aus Ellwangen/Jagst

Tübingen
2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	09.10.2020
Stellvertretender Dekan:	Prof. Dr. József Fortágh
1. Berichterstatter:	Prof. Dr. Markus Janzcyk
2. Berichterstatter:	Prof. Dr. Barbara Kaup
3. Berichterstatter:	Prof. Dr. Alexandra Bendixen

Acknowledgments

First, I would like to thank my advisor Prof. Dr. Markus Janczyk. Thank you for the trust you put in me, for your patience, your motivation, and the immense knowledge you shared with me. Although the distance between your office and mine tremendously increased when you moved to Bremen, your continuous support did not stop. You introduced me to the study of cognitive psychology in an enjoyable and challenging way. When it was necessary you pushed me forward and you also trusted me in times when even I myself did not believe that I should succeed. I am grateful that I could learn so much from you.

I am also very grateful to Prof. Dr. Barbara Kaup for coming onto my committee very last minute.

I also thank the members of the SFB 833 Project B2: Prof. Dr. Gerhard Jäger and Prof. Dr. Michael Franke. Furthermore, I want to thank Dr. Nadine Bade. Without being a formal member of the project you listened to this work in its many stages and were always ready with new ideas and suggestions. I am also thankful to Beate Starke for everything she does – without you, we would be lost.

Special thanks goes to the wonderful PhD candidates Dr. Moritz Durst, Dr. Ruben Ellinghaus, Fabiola Reiber, and Linda von Sobbe who devoted many hours of their time discussing everything from my data, to formal problems and ways of presenting the results with me. Although you were working in different research fields, you took the time to ask critical questions and you made interesting and fruitful suggestions. Thanks to those people who made working here a wonderful experience, especially: Prof. Dr. Rolf Ulrich, Dr. Karin Bausenhart, Dr. Daniel Bratzke, and Dr. Donna Bryce. Having you as colleagues made the time in the psychological department very enjoyable. Thank you very much.

I also want to thank our Hiwis Jan Bursik, Stella Faber, Philippa Hennessey, Carl Reichert, Sophie Renner, Lisa Schmidt, Kerstin Thieme, and Roswitha Vogel for their creative help while creating the items for the experiments, and the data collection. Special thanks go to Carolin Schonard who helped me with the statistical analysis of the mouse-tracking data. Thanks to Sara Wallochny, for finding and correcting many typos in the thesis.

Besides my wonderful colleges, I want to thank Ann for her patience and support during the whole project *dissertation*. Thank you for supporting me in more ways than you can imagine.

Research for this thesis was conducted within Project B2 of the Collaborative Research Centre 833 "The Construction of Meaning"

Contents

I Abstract

A presupposition is a phenomenon whereby speakers mark linguistically the information that is presupposed, rather than actually uttering it. The definite determiner (as in *the banana)* triggers the uniqueness-presupposition that there is a uniquely identifiable banana in the relevant discourse context. In contrast, the indefinite determiner (as in *a banana*) is associated with anti-uniqueness (that there are several bananas). The present dissertation investigates how this anti-uniqueness inference arises. Mouse-tracking studies (Study 1 and Study 4) provide evidence for the application of the Maximize Presupposition principle to the indefinite determiner. According to this principle the anti-uniqueness of the indefinite determiner results indirectly as an anti-presupposition from considering the uniqueness-presupposition of the definite determiner, which is then negated. This two-step processing results in increased processing difficulty for the indefinite determiner.

Furthermore, the present dissertation investigate what cognitive resources are involved during presupposition processing. To answer this question, the Psychological Refractory Period (PRP) approach is combined with a reading task to apply the locus of slack-logic (Study 2 and Study 3). The data provide evidence for a capacity-limited processing that has its locus within the central stage of processing and limited cognitive capacities are required during presupposition processing. Consequently, presupposition processing is not automatic and thus cannot run in parallel to other tasks. In addition, evidence for an immediate use of presuppositional information (Study 4 and Study 5) was gathered. In cases where the context explicitly falsifies the presuppositional content processing difficulties arise as early as the presuppositional content is fully known. The present results support a semantic view on presuppositions where the additional meaning component is encoded into the lexical entry of the trigger.

.

II Zusammenfassung

Eine Präsupposition ist eine Bedingung an den Kontext. Nur in Kontexten, die die Präsupposition beinhalten, kann ein Satz angemessen geäußert werden. Der definite Artikel (beispielsweise „die Banane“) löst die Einzigartigkeitspräsupposition aus, dass genau eine identifizierbare Banane im entsprechenden Diskurs existiert. Der indefinite Artikel (beispielsweise „eine Banane“) drückt eher Nicht-Einzigartigkeit aus, und deutet dadurch darauf hin, dass im Kontext mehrere Bananen existieren. Die vorliegende Dissertation beschäftigt sich mit der Frage, wie diese Nicht-Einzigartigkeitsinferenz entsteht. Mouse-tracking Studien (Studie 1 und Studie 4) liefern Evidenz für die Anwendung des Prinzips der *Präsuppositionsmaximierung* (engl. *Maximize Presupposition*). Nach diesem Prinzip entsteht die Nicht-Einzigartigkeit des indefiniten Artikels indirekt als Anti-Präsupposition nach vorheriger Betrachtung der Einzigartigkeitspräsupposition des definiten Artikels. Die zweistufige Verarbeitung des indefiniten Artikels beinhaltet als erstes die Verarbeitung der Einzigartigkeitspräsupposition des definiten Artikels und negiert diese im zweiten Verarbeitungsschritt. Durch diese zweistufige Verarbeitung entstehen erhöhte Verarbeitungskosten für den indefiniten Artikel.

Außerdem, untersucht die vorliegende Arbeit, welche kognitiven Kapazitäten während der Verarbeitung von Präsuppositionen beansprucht werden. Um diese Frage zu beantworten, wird das Paradigma Psychological Refractory Period (PRP) mit einem Lesezeitexperiment kombiniert um die Locus of slack-Logik anwenden zu können. Dieses Paradigma ermöglicht es herauszufinden, welche Verarbeitungsstufen beteiligt sind. Studie 2 und Studie 3 lieferten Evidenz für eine kapazitätslimitierte Verarbeitung von Präsuppositionen deren Ursprung in der zentralen Verarbeitungsstufe verortet wird. Es kann deshalb angenommen werden, dass die Verarbeitung von Präsuppositionen nicht automatisch stattfindet und dass diese Verarbeitung nicht parallel zu anderen Verarbeitungsprozessen ablaufen kann. Außerdem, konnte mit Studie

4 und Studie 5 Evidenz für eine schnelle Verarbeitung von Präsuppositionen geliefert werden. Wenn der Kontext die Präsupposition explizit falsifiziert, führt dies zu Verarbeitungsschwierigkeiten, die auftreten, sobald die Präsupposition dem Hörer bekannt ist, allerdings bereits vor dem Ende des Satzes. Die präsentierten Ergebnisse unterstützen einen semantischen Ansatz, wonach die zusätzliche Bedeutung, die durch Präsuppositionen ausgerückt wird im lexikalischen Eintrag des Präsuppositionsauslösers verankert ist.

III Publications

III.I Enclosed publications and author contributions

Study 1

Schneider, C., Schonard, C., Franke, M., Jäger, G., & Janczyk, M. (2019). Pragmatic processing: An investigation of the (anti-)presuppositions of determiners using mouse-tracking. *Cognition, 193,* 104024.

Author	**Author position**	**Scientific ideas %**	**Data generation %**	**Analysis & interpretation %**	**Paper writing %**
Schneider, C.	1	30%	65%	40%	50%
Schonard, C.	2	0%	0%	30%	5%
Franke, M.	3	30%	5%	20%	5%
Jäger, G.	4	10%	0%	10%	5%
Janczyk, M.	5	30%	30%	30%	35%
Title of paper:		Pragmatic processing: An investigation of the (anti-) presuppositions of determiners using mouse-tracking.			
Status in publication process:		Published			

Study 2

Schneider, C., & Janczyk, M. (accepted). Capacity limitations of processing presuppositions triggered by determiners. *Acta Psychologica.*

Author	**Author position**	**Scientific ideas %**	**Data generation %**	**Analysis & interpretation %**	**Paper writing %**
Schneider, C.	1	40%	75%	65%	75%
Janczyk, M.	2	60%	25%	35%	25%
Title of paper:		Capacity limitations of presupposition processing.			
Status in publication process:		accepted for publication on 24.07.2020			

Study 3

Schneider, C., Bade, N., & Janczyk, M. (2020). Is immediate processing of presupposition triggers automatic or capacity-limited? A combination of the PRP approach with a self-paced reading task. *Journal of Psycholinguistic Research 49,* 247-273.

Author	**Author position**	**Scientific ideas %**	**Data generation %**	**Analysis & interpretation %**	**Paper writing %**
Schneider, C.	1	40%	75%	45%	65%
Bade, N.	2	20%	0%	20%	10%
Janczyk, M.	3	40%	25%	35%	25%
Title of paper:		Is immediate processing of presupposition triggers automatic or capacity-limited? A combination of the PRP approach with a self-paced reading task.			
Status in publication process:		Published			

Study 4

Schneider, C. Bade, N., Franke, M., & Janczyk, M. (2020). Presuppositions of determiners are immediately used to disambiguate utterance meaning. *Psychological Research*, https://doi.org/10.1007/s00426-020-01302-7.

Author	**Author position**	**Scientific ideas %**	**Data generation %**	**Analysis & interpretation %**	**Paper writing %**
Schneider, C.	1	25%	70%	55%	40%
Bade, N.	2	25%	0%	5%	20%
Franke, M.	3	25%	5%	10%	10%
Janczyk, M.	4	25%	25%	30%	30%
Title of paper:		Presuppositions of determiners are immediately used to disambiguate utterance meaning			
Status in publication process:		Published			

Study 5

Schneider, C., Janczyk, M., & Bade, N. (under review). Verifying pragmatic content in context: An experimental comparison of presuppositions of *again* and the definite determiner with scalar implicatures. *Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data.*

Author	**Author position**	**Scientific ideas %**	**Data generation %**	**Analysis & interpretation %**	**Paper writing %**
Schneider, C.	1	40%	75%	50%	60%
Janczyk, M.	2	30%	25%	20%	20%
Bade, N.	3	30%	0%	30%	20%
Title of paper:		Verifying pragmatic content in context: An experimental comparison of presuppositions of *again* and the definite determiner with scalar implicatures			
Status in publication process:		under review			

III.II Conference contributions and invited talks

Schneider, C., & Janczyk, M. (2020). Is processing of presuppositions triggered by determiners an automatic or capacity-limited process? Vortrag, 62. Tagung experimentell arbeitender Psychologen (TeaP), 22-25th März 2020, Jena [conference was cancelled due to Corona pandemia].

Schneider, C., Janczyk, M., & Bade, N. (2020). Presupposition verification: A comparison of definite determiners and again. Poster, Linguistic Evidence 2020, 13-15th February 2020, Tübingen, Germany.

Schneider, C., & Janczyk, M. (2020). Capacity Limitations of presupposition processing. Talk, 1st GK Doctoral Symposium on Cognitive Science, 24th January 2020, Tübingen, Germany.

Schneider, C., Bade, N., & Janczyk, M. (2019). Is immediate processing of presupposition triggers automatic or capacity-limited? Poster, 60th Psychonomic Society Annual Meeting (Psychonomics), 14th November 2019, Montreal, Canada.

Schneider, C., Franke, M., Schonard, C., Jäger, G., & Janczyk, M. (2018). *Pragmatic Processing: An Experimental Investigation of Presuppositions Using Mouse-Tracking.* Poster, 59th Psychonomic Society Annual Meeting (Psychonomics), 15th November 2018, New Orleans, USA.

Schneider, C., Franke, M., Schonard, C., Jäger, G., & Janczyk, M. (2018). *Pragmatic Processing: An Experimental Investigation of Presuppositions Triggered by Definite Determiners.* Poster, Workshop für Doktorand*innen der allgemeinen Psychologie (A-Dok), 15th Juli 2018, Mainz.

Schneider, C., Franke, M., Schonard, C., Jäger, G., & Janczyk, M. (2018). *Pragmatic Processing: An Experimental Investigation of Presuppositions Triggered by Definite Articles.* Vortrag, 60. Tagung experimentell arbeitender Psychologen (TeaP), Session *Language,* 13th März 2018, Marburg.

1. Introduction

Language and communication are omnipresent in everyday life and often speakers communicate more than they actually say. How listeners grasp the overall meaning of an utterance in a certain context is an important question in the study of language meaning. Besides the syntax, which is the appropriate structure of the sentence, and the correct semantics dealing with the intended meaning of the individual words, the pragmatic component of a sentence in the respective context is important for successful communication. In a situation with one pen and two pencils on the table it is absolutely okay to utter: “Please hand me the pen.” This request becomes odd, when the context is slightly modified and there are suddenly two pens on the table. This minimal change in the contextual setting causes a failure in communication. The listener is unable to fulfill the request because it is unclear which pen the speaker refers to. The present example shows that in situations where the presupposition of the definite determiner (which is existence and uniqueness (see e.g., Strawson, 1950)) is not satisfied, communication fails. To resolve this problem, the speaker is expected to use the indefinite determiner as in “Please hand me a pen”. While the definite determiner presupposes existence and uniqueness, the indefinite determiner is associated with anti-uniqueness. How those additional meaning components of the definite and indefinite determiner arise and which underlying processes are involved in the processing of such presupposition triggers is the question of the current dissertation. Although presuppositions have been a vital topic in the semantic and pragmatic literature throughout the last decades (for an overview see, e.g., Beaver & Geurts, 2012), there are many open questions, concerning the processing and interpretation of the definite and indefinite determiner. To gather experimental evidence for the account *Maximize presupposition* proposed by Heim (1991), I use mouse-tracking experiments and combine linguistic methods with a paradigm from cognitive psychology. The data provide novel insights in the underlying processes of presupposition processing from a linguistic and cognitive psychology perspective. Before presenting the experimental results, I will introduce the theoretical background of presuppositions from a semantic and pragmatic perspective in Chapter 2. In Chapter 3 experimental methods with relevance for the present dissertation are explained and four important experimental results are discussed. In Chapter 4, I will make a brief excursus to the field of cognitive psychology to describe the Psychological Refractory Period (PRP) approach and the locus of slack-logic which was used in my own experiments to form hypotheses about the underlying processes of presupposition processing. Having covered all necessary theoretical background, I will lead to the research questions of the dissertation in Chapter 5. The conducted experiments and their results will be discussed in Chapter 6. In Chapter 7, I will discuss the results and I will show what implications those results have for the theories of presuppositions. Answers to the research questions are provided in Chapter 8.

2. Theoretical background of presuppositions

In this Chapter, I will provide the theoretical background for presuppositions and introduce two major theoretical approaches. Presuppositions are ubiquitous in everyday life and they are a vital topic in the semantic and pragmatic literature (e.g., Beaver & Geurts, 2012, for an overview). To get a deeper understanding of this phenomenon, I want to give an overview on the general properties of presuppositions and first of all, answer the question of how a presupposition can be distinguished from the asserted meaning of a sentence.

As a first approximation, presuppositions have been distinguished from assertions by the following test called „Hey, wait a minute "-test (e.g., Von Fintel, 2004):

(1) Anna's dog likes to go for a run.

(2) Hey wait a minute, I didn't know that Anna has a dog.

The sentence in (1) contains the presupposition that Anna owns a dog. In contexts where this additional meaning component is not clear to the addressee the presupposition can be challenged with the response in (2). This test cannot be applied to an assertion like "Dogs like to play with tennis balls" because it would be odd to respond "Hey wait a minute, I didn't know that dogs like to play with tennis balls." The „Hey, wait a minute "-test can only be applied to presuppositions, but not to assertions and thus can be used to distinguish those two meaning components.

Besides this test there is another feature that distinguishes presuppositions from assertions and from other pragmatic phenomena like implicatures. Presuppositions survive embedding under certain operators such as negation (e.g., Chemla, 2009), questions (e.g., Karttunen, 1973), conditionals (e.g., Karttunen, 1973), and modals (e.g., Langendoen & Savin, 1971). This can be seen in the example in (3). All of the example sentences in (3a-e) presuppose that Isabel used to smoke.

(3) a. Isabel stopped smoking. (simple sentence)

b. It is not the case that Isabel stopped smoking. (negation)

c. Has Isabel stopped smoking? (question)

d. If Isabel stopped smoking, she would feel much fitter. (conditional)

e. Isabel might stop smoking. (modal)

This does not hold for implicatures or assertions as can be seen in the examples in (4-5). While in (4a) some, but not all of Anna's friends play soccer, the negation of the sentence changes the meaning of the sentence. These phenomena of presuppositions surviving under negation and questions is called *projection*.

(4) a. Some of Anna's friends play soccer. (simple sentence)

→implicature: not all of Anna's friends play soccer

b. It is not the case that some of Anna's friends play soccer. (negation)

→ no implicature

c. Do some of Anna's friends play soccer? (question)

→ no implicature

(5) a. Dogs like to play with balls. (assertion)

b. Dogs don't like to play with balls. (assertion)

A second question that I seek to answer is: What causes presuppositions? The literature refers to presupposition triggers. Frege (1892) was the first one who used the term presupposition in his work on definite descriptions. This was the birth of the term presupposition trigger. Frege's work lead to a famous argument between Russel (1905) and Strawson (1950) about meaning and use of determiners. Since then, many linguistic expressions and syntactic constructions have been argued to cause presuppositions to arise. In the following, I want to provide a list of so-called presupposition triggers following work by Beaver and Geurts (2012):

- **Definite descriptions** (Strawson, 1950)

  The national soccer team of Togo scored a goal.

  Presupposition**:** Togo has a (unique) national soccer team.

- **Iteratives** (von Stechow, 1996)

  Anna scored a goal again.

  Presupposition**:** Anna scored a goal before.

- **Factives** (Kiparsky & Kiparsky, 1970)

  Leopold knows that Anna plays soccer.

  Presupposition: Anna plays soccer.

- **Aspectual verbs** (Simons, 2001)

  Anton stopped playing tennis.

  Presupposition: Anton used to play tennis.

- **Manner adverbs** (Abbott, 2000)

  Anna ran quickly to the coach.

  Presupposition**:** Anna ran to the coach.

- **Temporal clauses headed by *before, since, after*, etc.** (Beaver & Condoravdi, 2003)

  Anna talked to the referee before she took a free kick.

  Presupposition: Anna took a free kick.

- **Sortally restricted predicates of various categories** (*e.g., bachelor*) (Thomason, 1972)

  Julius is a bachelor.

  Presupposition: Julius is an adult male.

- **Cleft sentences** (Delin, 1995)

  It was Peter who got a red card.

  Presupposition: Somebody got a red card.

- **Quantifiers** (Cooper, 1983)

  Anna passed to every player on the pitch.

  Presupposition: There are players on the pitch.

- **Names** (van der Sandt, 1992)

  The coach is Theresa Merk.

  Presupposition: Theresa Merk exists.

- **Intonation (e.g., focus, contrast)** (Geurts & van der Sandt, 2004)

  SHE scored a goal.

  Presupposition: Somebody scored a goal.

As can be seen, the list of expressions and constructions that trigger presuppositions has grown tremendously and they comprise a very heterogeneous bunch of lexical and syntactic items. It seems therefore necessary to turn away from one theory that can treat all presupposition triggers the same. Work by Simons (2001), Abusch (2002), and Abrusán (2011) suggests to categorize the triggers and then evaluate different theories.

2.1 Semantic presuppositions

The central idea from a semantic perspective is thus that all presupposition triggers have the respective presupposition encoded into their lexical entry. The example in (6) can only be uttered felicitously in contexts that entail that Anna scored before yesterday. Otherwise the sentence is infelicitous.

(6) Yesterday, Anna scored again.

The relevant lexical entry for the trigger *again* can be seen in (7). *again* is one of the triggers that does not contribute anything to the assertive meaning of a sentence. The contribution is only on a presuppositional level.

(7) $[[again]] = \lambda w.\ \lambda t'.\ \lambda P_{<s;<i;t>>}.\lambda t'': t' < t'' \ \&\ P(w)(t')P(w)(t'')$

As a second argument *again* takes a time variable because *again* calls for a specific time interval in which the relevant proposition holds true and not just for any time interval. There seems to be a parallel between the behavior of *again* and *too*. Both do not contribute to the assertive meaning of a sentence. For the lexical entry, I followed a suggestion of Beck (2007) in (8).

(8) $[[\text{ too }]] = \lambda w.\ \lambda P<s;<e;t>>.\lambda x\text{: } \exists y[y \neq x \ \&\ P(w)(y)].P(w)(x)$

again and *too* are presupposition triggers that do not make an assertoric contribution and they are referential.

Another group of presupposition trigger is still referential, but also makes a contribution to the assertion. The example in (9) shows that strange inferences can arise when taking the conventional view on the trigger *stop*.

(9) John is cooking.

He will stop (cooking) when tomorrow's football game starts.

In this case, *stop* introduces a presupposition which requires that there is an event which lasts until just before the reference time of the sentence. The example in (9) however, makes clear that this event (or the running time of the event) is specific rather than existential. In (10) I provide the lexical entry for *stop*. Similar to *again* and *too*, *stop* also takes a variable as its second argument which will remain free in the course of the compositional interpretation. On the other hand, *stop* makes a meaningful contribution to the assertion.

(10) $[[\text{stop}]] = \lambda w.\ \lambda t'.\ \lambda P<s;<i;t>>.\lambda x.\ \lambda t\text{: } t' \infty\ t \ \&\ P(w)(t')(x).\neg P(w)(t)(x)$

The central assumption according to the semantic view is that the presupposition is encoded directly in the trigger and therefore, the additional meaning component is already available while reading the trigger. Following this approach (see Frege, 1892; Heim, 1982;

Russel, 1905), a presupposition is associated directly with the trigger word and thus immediately leads to awareness of the importance of context. This leads to the prediction that immediate processing of presuppositions starts already at the trigger word and evokes processing costs early before finishing the sentence (e.g., Bade & Schwarz, 2019b; Burkhardt, 2006; Kirsten et al., 2014; Tiemann et al., 2011).

2.2 Pragmatic presuppositions

There is an alternative theoretical perspective on presuppositions which takes a more pragmatic approach (Levinson, 1983; Simons, 2001; Stalnaker, 1973). It assumes that presuppositions are not semantically encoded but are pragmatic, und thus arise through pragmatic reasoning, that is, only play a role after the sentence's main meaning is computed and its integration into context is considered. According to this procedure, the trigger is not processed immediately, but only at the end of the sentence after the assertion.

One observation by Simons (2001) that supports this view is *contextual defeasibility* of some presupposition triggers like change of state predicates and factives.

(11) I notice that you keep chewing on your pencil. Have you recently stopped smoking?

The example in (11) is taken from Geurts (1994) and illustrates that not all presuppositions seem to be inherently semantic. Usually, presuppositions survive questions. However, the example in (11) shows that the presupposition of *stopped* seems to "disappear" due to the explicit ignorance context that is created via expressing the previous sentence. In such a context the addressee knows that the speaker is ignorant about the proposition that would normally arise as a presupposition (in this case the smoking habits of the addressee). The addressee is aware that the speaker is ignorant about the current or prior smoking habits of her and thus cannot assume that she recently used to smoke. The utterance is interpreted as a non-presupposing question asking whether the addressee has undergone the relevant change from being a smoker to not being one ("Is it the case that you have recently been a smoker and have recently ceased to be one?"). The non-presupposing interpretation is preferred over the presupposing one. In explicit ignorance contexts the presupposition of change of state verbs fail to arise. When looking at the example in (12), one can see that this does not hold for presuppositions triggered by *again*.

(12) # I don't know if Anna scored before, but perhaps she's scoring again.

The ability of projection of the presupposition trigger *again* is still there in example (12), although the first part of the sentence asserts explicit ignorance. This causes the oddness of the sentence. When looking at scalar implicatures as in (13) one can see that the presupposition behaves similarly to implicatures in certain environments. In (13) the scalar implicature (that John has three children and not more) is canceled by the second conjunct.

(13) John has three children and may have more for all I know.

Simon (2001) observed that some presuppositions behave similarly and thus takes this as an argument for their pragmatic status. Another parallelism between presuppositions and implicatures that Simons observed is *nondetachability.* The examples in (15a-d) illustrate that implicatures do not arise because of specific lexical expressions, but they are caused by pragmatic reasoning. All the examples in (15) convey the implicature that Sonja does not want to go out for a drink.

(14) Tom: Do you want to go out for a drink?

(15) a. Sonja: I have to finish writing my paper.

b. Sonja: I need to finish my paper.

c. Sonja: My paper needs to get finished tonight.

d. Sonja: I have to work on my paper.

By definition conversational implicatures are not conventionally associated with any expression. According to Simons (2001) the example in (16) illustrates that the same holds for change of state verbs and factives because the presupposition is not tied to a single lexical expression, but to the content expressed in the sentence. The synonyms in (16) give rise to the same presupposition, namely that Tom was a fool. That synonyms give rise to the same presupposition strongly suggests that presuppositions of change of state verbs and of factives are attached to the content and not to the lexical item itself.

(16) Tom didn't realize / come to know / become aware of that he was a fool.

Contextual defeasibility and *nondetachability* are observations that brings Simons (2001) to the conclusion that presuppositions triggered by factives and change of state verbs have a conversational source and thus are pragmatic in nature.

Further evidence for a pragmatic source of presuppositions comes from Abrusán (2011). According to her analysis, factive verbs, change of state predicates, emotives, and achievement verbs trigger presuppositions which are determined by the cognitive language system in a bottom-up process. This process determines which entailments of a sentence are not about the same time interval as the main clause predicate. Abrusán (2011) argues that it is precisely these entailments which end up as being presupposed.

The crucial point in this assumption is that this process can only operate over sentence entailments and therefore the presupposition of a sentence can only be determined after the sentence's truth conditions are established. For the processing of presuppositions this means that presupposition processing can only succeed the processing of the asserted meaning. An immediate processing of presuppositions already on the trigger is therefore impossible according to this pragmatic view.

In the semantic and pragmatic literature, there is still a controversial debate about how presupposition arise ("the triggering problem"). One question that has arisen is whether there actually is the need for a separate concept of presuppositions or whether the issue is better understood in terms of what is at-issue or raises attention versus what is non-at-issue/in the background (Abrusán, 2011; Tonhauser, Beaver, & Degen, 2018). Additionally, there is still an ongoing debate of how presupposition triggers can be classified and whether it is possible to make uniform predictions for all triggers or whether they have to be treated separately.

2.3 (Anti-) uniqueness expectations of (in-) definite determiners

The definite determiner is a classic example for a presupposition trigger. It introduces the presupposition of existence and uniqueness as can be seen in (17a). There exists a unique individual that has the property of the noun the definite determiner combines with, in this case the property of being the mother of the soccer coach.

(17) a. The mother of the soccer coach arrived at the stadium.

b. # A mother of the soccer coach arrived at the stadium.

c. A brother of the soccer coach arrived at the stadium.

The theoretical literature claims that definite determiners have to be used if their presupposition is fulfilled by the context. (17b) shows that it is infelicitous to use the indefinite determiner in those contexts. As can be seen in (17c) using the indefinite determiner gives rise to the assumption that there exists more than one brother and thus the indefinite determiner is associated with anti-uniqueness.

There are various theories explaining this effect. Following Kratzer (2005), the indefinite determiner comes with its own presupposition of anti-uniqueness (see also the discussion in Heim, 1991, 2011). Consequently, both determiners introduce their own restrictions on what is regarded as an appropriate context and thus they should not differ in processing.

However, this attempt cannot explain why the indefinite determiner is inappropriate in contexts where it is common knowledge that the uniqueness-presupposition is met as in (17a).

Furthermore, Heim (1991) observed that the indefinite determiner in (18) can be uttered without being certain that there is more than one 20-ft-long catfish. The utterance is felicitous as long as the speaker is not sure that there is exactly one such catfish.

(18) Robert caught a 20-ft-long catfish.

To account for the anti-uniqueness effect in (17b and c) and the felicity of (18), Heim (1991) proposed to add another principle to Grice’s (1975) conversational maxims: Maximize Presupposition. This principle says: Presuppose as much as possible! (see also Chemla, 2009; Percus, 2006; Sauerland, 2008; Schlenker, 2012). It requires the speaker to always use the felicitous sentence with the strongest presupposition among a set of alternatives, as long as the speaker knows that these presuppositions are fulfilled. This can explain the oddness of the indefinite determiner in the example in (17b). The sentence in (17a) is an alternative, because the only difference is the used determiner. The definite determiner in (17a) introduces more presuppositions that are true in the context (since we know that people have a unique mother) and thus is the presuppositionally stronger alternative. When the speaker utters the presupositionally weaker sentence (17b), the listener assumes that the speaker must believe the presupposition of the stronger alternative to be false. This assumption is based on two prerequisites: First, the speaker obeys the conversational maxims including the Maximize Presupposition principle and second, the speaker is cooperative. Is this the case, the hearer can assume that if the speaker believed the presupposition of (17a) to be true s/he would have uttered this sentence because it triggers more presuppositions (existence and uniqueness). Since the speaker did not do that, s/he must believe that those presuppositions do not hold. This leads to the belief that the soccer coach has more than one mother, which is contrary to common knowledge and thus results in the oddness of (17b).

As pointed out by the explanation of the oddness that arises in example (17b), the anti-uniqueness of the indefinite determiner arises by considering the stronger alternative of the definite determiner and a subsequent negation of its presupposition. This leads to the conclusion that the inferences that arise based on Maximize Presupposition are not presuppositions proper. Rather, they are referred to as anti-presuppositions in the literature (Percus, 2006). Following this approach, deriving the anti-uniqueness inference of the indefinite determiner is a two-step process. This process involves an initial consideration of the stronger alternative (the uniqueness-presupposition of the definite determiner) and its subsequent negation. Consequently, the processing of the indefinite determiner should be more complex than the processing of the definite determiner.

An alternative theory regards the indefinite determiner as a scalar implicature because it is in competition with other quantificational terms like *every* or *all* (Chierchia, Fox, & Spector, 2012; Grønn & Sæbø, 2012). The indefinite determiner and the quantificational terms can be ordered on a lexical scale (Horn, 1972). A scalar implicature arises when the speaker choses a weaker item from that scale (in this case the indefinite determiner). Assuming that the speaker obeys the conversational maxims and that s/he is cooperative, all items that are higher in the scale (and are thus stronger alternatives) become negated. For example, the implicature of “A girl scored” is that “Not all girls scored”. The competition of alternatives is on the level of assertion (not on the presuppositional level according to Maximize Presupposition). According to Magri (2009), the anti-uniqueness inference of the indefinite determiner is an implicature and follows from the general mechanism of exhaustification over alternatives (see also Schlenker, 2012; Singh, 2011). Following this approach, a faster processing of the indefinite compared to the definite determiner is predicted because implicatures are assumed to be processed faster than presuppositions (see Bill, Romoli, & Schwarz, 2018).

In my own experimental investigation, I will focus on the anti-uniqueness account based on the Maximize Presupposition principle. Only Study 4 took the two other explanations into consideration.

3. Important experimental results

In the following section, I focus on four different important results regarding presupposition processing. I will briefly introduce general experimental methods and then elaborate on the results. Briefly, the four results of special interest for the present dissertation are:

(I) Presupposition triggers induce **additional processing costs** compared to neutral words. Following a semantic approach, the presupposition is encoded in the lexical entry of the trigger itself. This additional meaning component has to be processed and this process causes processing costs. Therefore, the processing of presupposition triggers is more difficult than the processing of non-trigger words. Furthermore, due to the direct encoding in the trigger itself, processing of **presuppositions starts immediately on the trigger.**

Presuppositions vary according to the present context. When the context supports the presupposition, a verification process has been successful. This verification process is started as soon as the presupposition is fully known. A failure of the verification process leads to additional processing costs, for example, in case of definite determiners a memory search for a corresponding referent has to be terminated without success. **(II) Presupposition failure is more demanding than verification.**

Frege's (1892) analysis of the definite determiner suggests that it comes with the presuppositions of existence and uniqueness, while indefinite determiners have none. As mentioned in section 2.3. there are different theoretical approaches towards the (anti-) uniqueness of the (in)definite determiner and their processing. In the following, experimental evidence is provided that shows a difference in processing between the definite and the indefinite determiner. **(III) The processing of the definite and the indefinite determiner differs.**

If the context neither supports the presupposition nor directly falsifies it, a rescue strategy is applied. In order to make sense of the sentence, the context gets updated with the information conveyed by the presupposition. This strategy is called accommodation. In case accommodation takes place, the process is even more demanding than the verification of a presupposition or the failure of this process. **(IV) Accommodation of presuppositions is even more demanding than verification or falsification.**

3.1 Experimental methods

In this overview, I will focus on four methodological approaches. In the following paragraph, I will briefly introduce the general approach of the different experimental methods and I will describe particular experiments providing evidence for the mentioned effects.

3.1.1 Acceptability ratings

Acceptability ratings of sentences in a given context offer a straightforward way to measure a core feature of presuppositions, namely that people take their content for granted. Therefore, participants have to read a context sentence and a test sentence and have to rate the appropriateness of the test sentence according to the presented context. This procedure is easy to carry out without much technical effort. It is a useful way to find out more about the relationship between presupposition triggers and context. It is often used and combined with other methods to get insight into the way participants understand the test sentences and to verify that participants interpret the sentences in the intended way. However, one can only get data about the final interpretation of a sentence (off-line) and no measures that can give insight in the time course of processing (on-line).

3.1.2 Reaction times

Reaction time measures are frequently used to investigate comprehension processes. In psycholinguistics such measures are used to examine the relative time-course of a process. According to Rayner and Clifton (2002, p. 3), **reaction time** (RT) is defined as “the interval between the presentation of a stimulus and the onset of the subject’s subsequent response”. Nowadays, RTs are measured with a high degree of precision (e.g., in milliseconds). When examining the reader’s comprehension of sentences, characteristics of a text are manipulated and participants are instructed to read the sentence while reading times are recorded. In those experiments, participants are instructed to read a sentence or a whole paragraph while the elapsed time is recorded. With this method, researchers can infer selected attributes of comprehension processes.

Sometimes overall **reading times** of a sentence are too imprecise, because one is interested in the time a subject takes to read a particular segment of a text. Then **self-paced reading** is applied. The self-paced reading paradigm goes back to Just, Carpenter, and Woolley (1982). Participants are instructed to read sentences and they can control via button presses the exposure duration of single segments of the sentence (the experimenter defines those segments in advance according to the experimental goal). The latencies of the button presses depend on the properties of the words being read and correlate with the time course of the cognitive processes during reading and text comprehension. The so-called moving-window technique is widely used in psycholinguistics and was identified to most closely resemble natural reading while measuring reading times for single words/sections is possible. In a moving-window self-paced reading experiment, the test sentence is first displayed as a series of dashes on the screen, with each dash representing a word in the sentence. The length of the dashes correlates with the word length. When the participant first presses a button, the first word appears on the screen replacing the corresponding dash. Subsequent button pressing causes the previous word to disappear and to be replaced by a dash while the current word appears on the screen. Only one word (or phrase) is visible at any given time, thus creating the impression of a moving window of words on the screen. Contrary to natural reading, regressions to previous words are impossible. Self-paced reading times help to get insight into the time course of processing of the interpretation of presuppositions, but on a coarse-grained level. In general, reading times in self-paced reading are slower than those in natural reading (according to Rayner & Clifton, 2002, about the double time), which can facilitate the development of alternative comprehension strategies compared to natural reading.

3.1.3 Eye-tracking

Eye-tracking can be a more fine-grained method to study online processing, and is often used in combination with self-paced reading. It involves tracking of eye-movements and it is a non-invasive, sensitive tool that quantifies and measures eye-movements to describe an individual's cognitive state. Findings by Rayner and Clifton (2002) reflect that the point at which readers look and the time they spend looking at a specific point is directly related to the difficulty of cognitive processing. Eye-movement data can be used to investigate lexical ambiguity, morphological complexity, or discourse processing.

3.1.4 Electroencephalogram (EEG)

In the past 40 years, electrophysiological measures have been developed to investigate cognitive processes and to get insight into the time course of cognitive activity within the brain. The fact that transmission of information in the brain involves the flow of ions is used to measure electrical activities of the brain. When ions flow across a neural membrane, a voltage field around each active neuron results. During invasive intracranial recordings the activity of a single neuron can be monitored. Furthermore, the electrical fields around neighboring neurons also produce a field that can be detected further away, even as far away as the scalp.

The voltage fluctuation that is produced by a large population of neurons can be recorded by scalp electrodes. The recordings trace the voltage across time which is known as electroencephalogram (EEG). Any given tracing reflects the differences in electric potential (i.e., voltage) between two recording sites. With developing computer technologies, it became possible to estimate the activity time-locked to a certain point, for example, the stimulus onset. Therefore, the electrical potential over a number of trails is averaged, the so-called **event-related potential** (**ERP**). A major advantage is the more direct examination of the time course of language comprehension within the brain itself. A high amount of data makes it possible to track processes throughout the course of a sentence within a single subject. With appropriate stimulus material and coding, one can examine more than one issue within a single study. On the other hand, it is hard to identify what the direct source of the ERP activity is and which cognitive process is responsible. Furthermore, the interpretation of ERP can be complicated because multiple ERP components can occur in the same latency range. Another problem occurs during recording the EEG. Only little motor activity for the subject is possible because eye movements, activity of facial muscles, or tongue movements produce their own electrical artifacts that influence the EEG recording. This means that participants should stay relatively still during the recording which makes it more unnatural.

Having introduced general experimental methods that are used to investigate the processing of presuppositions, I will present four major experimental results that are of interest for the present dissertation in more detail.

3.2 Effect 1: Presupposition processing starts immediately on the trigger and is more difficult compared to neutral words

The processing of presuppositions starts immediately on the word triggering the presupposition. This leads to additional processing costs for trigger words compared to neutral words.

3.2.1 Reading times

Tiemann et al. (2011) investigated the processing of presuppositions in a series of three self-paced reading experiments combined with acceptability ratings. They analyzed the processing of presuppositions induced by different triggers (German *wieder* ('again'), *auch* ('also'), *aufhören* ('stop'), *wissen* ('know'), and *definites* in the shape of possessive noun phrases). In their first experiment, the focus was on the trigger itself and reading times of test sentences including a presupposition trigger (e.g., (19a)) were compared to grammatical sentences without a presupposition (e.g., (19b)), and with ungrammatical sentences (e.g., (19c)).

(19) **Context**: Tina ist mit einer guten Freundin shoppen.
Tina is shopping with a good friend.

a. Sie kauft wieder rote Handschuhe.
She buys red gloves again.

b. Sie kauft heute rote Handschuhe.
She buys red gloves today.

c. *Sie kauft freundlich rote Handschuhe.
She buys red gloves friendly.

The grammatical sentences without a presupposition were rated best, the sentences including the presupposition trigger were rated slightly worse, pointing to the fact that accommodation has to take place and the unacceptable sentences were rated worst. The reading times revealed that for the position of the trigger (or the corresponding neutral/unacceptable word), the trigger condition induced the longest reading times followed by the neutral condition and the unacceptable condition evoked the shortest reading times. This suggests that the trigger needs more attention because it alerts the reader to look back at the preceding context. The early effects support the idea of an immediate processing of presuppositions starting on the trigger itself. Unfortunately, reading times were not analyzed for individual presupposition triggers, as there were not enough items for each trigger to allow for strong conclusions. It is thus unclear whether the same pattern holds for all triggers.

Schwarz (2007) also reports self-paced reading results for the trigger *also* for both German and English. The presupposition was either satisfied or unsatisfied within the presented sentence. The data revealed longer reading times on the region containing *also* in the latter case.

3.2.2 Eye-tracking

Bade and Schwarz (2019b) provide further evidence for an immediate processing of presuppositions with a visual-world eye-tracking study in a picture selection task. Participants were instructed to identify an individual called Benjamin after they heard an auditorily presented test sentence like "The/A shirt in Benjamin's closet is blue." They had to select one of three pictures all of which depicted a boy with a closet. In one of the pictures the closet contained three shirts, one of which was blue (non-unique condition). Another picture showed only one blue shirt (unique condition), and in the third picture (distractor condition) no shirt was depicted. The results show that participants looked at the respective target picture (picture with a single shirt for the definite determiner, and picture with multiple shirts for the indefinite determiner) very early on upon hearing the noun. This suggests that the information about uniqueness or anti-uniqueness encoded in the determiners was used rapidly for interpretation of the test sentence, already while hearing the determiner itself.

Further evidence for an immediate integration with the discourse context was provided by Schwarz and Tiemann (2012). Sentences in German with the trigger *wieder* ('again'), in contexts that either are or are not consistent with its presupposition showed slowdowns in the earliest fixation measures (e.g., first fixation duration, first pass regression proportion) of the verb that immediately follows *again*, supporting the idea of immediate processing starting on the trigger.

3.2.3 EEG

Further support for an immediate processing of presupposition triggers comes from EEG studies by Kirsten et al. (2014). They investigated the processing of definite and indefinite determiners in two types of context sentences (see (20 a) and b)): either introducing a single object (e.g., one polar bear) or multiple objects (e.g., some polar bears). The test sentences were similar except for used determiner (the/a, see (21a) or b)) and they were presented either in a matching condition where the context sentence introduced the noun phrase with a quantifier “ein/e” (Engl.: “a”) or in the mismatching condition which contained a quantifier such as “einige” (Engl.: “some”) or “viele” (Engl.: “many”).

(20) a. Antje war gestern im Zoo in Düsseldorf und besuchte einen Eisbären im Bärengehege.
Antje visited the Düsseldorf zoo yesterday and saw a polar bear in the bear enclosure.

b. Antje war gestern im Zoo in Düsseldorf und besuchte einige Eisbären im Bärengehege.
Antje visited the Düsseldorf zoo yesterday and saw some polar bears in the bear enclosure.

(21) a. Antje beobachtete, dass der Eisbär sehr aggressiv war.
Antje noticed that <u>the</u> polar bear was very aggressive.

b. Antje beobachtete, dass <u>ein</u> Eisbär sehr aggressiv war.
Antje noticed that <u>a</u> polar bear was very aggressive.

The data revealed that participants recognized the mismatching condition already when reading the determiner. For the indefinite determiner, the mismatching effect became visible within the N400 and the P600 time window. The N400 was interpreted as a semantic mismatch (Kutas, Van Petten, & Kluender, 2006) and the P600 as an index for a subsequent reanalysis process (Kuperberg, 2007; Van de Meerendonk, Kolk, Vissers, & Chwilla, 2008). Due to the early effects, Kirsten et al. (2014) concluded that presupposition processing begins as soon as the presupposition trigger is perceived.

Furthermore, general differences between the definite and the indefinite determiner were observed, namely that the definite determiner evoked an enhanced negativity compared to the indefinite. Burkhardt's (2006) ERP study further supports the idea of early processing of presuppositions by revealing an N400 effect on the trigger position when the existence presupposition of the definite determiner was not satisfied.

3.3 Effect 2: Presupposition failure is costly

Failure of the verification process leads to additional processing cost: infelicitous use is more difficult than felicitous use.

3.3.1 Reading times

Although Altmann and Steedman (1988) wanted to test Frazier's (1978) *Minimal Attachment Hypothesis*, their results turn out to be very relevant for a theory of presupposition processing. In a self-paced reading study, they investigated how sentences with a definite determiner are processed if the uniqueness-presupposition is not satisfied and they observed effects before the end of the sentence. Test sentences were presented in two different contexts. Context 1 (see (22)) introduced two candidates as referent (a safe with a new lock and a safe with an old lock) while Context 2 (see (23)) introduced exactly one candidate as referent (a safe with a new lock).

(22) Context 1: A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was <u>a safe with a new lock</u> and a safe with an old lock.

(23) Context 2: A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was <u>a safe with a new lock</u> and a strongbox with an old lock.

In the test sentences in (24), the prepositional phrase *with the new lock* modified *the safe* and therefore the uniqueness-presupposition is satisfied even if the sentence is presented in a context like (21), whereas this is not the case for the test sentence in (25). The test sentence in (24) can only refer to the single safe mentioned in the context in (23) otherwise (in context (22)) the reader does not know which safe was blown open.

(24) The burglar / blew open / the safe / with the new lock / and made off / with the loot.

(25) The burglar / blew open / the safe / with the dynamite / and made off / with the loot.

The results of this experiment revealed that reading times differed during the disambiguating region (i.e., the prepositional phrase *with the new lock* or *with the dynamite*). Test sentences with an unsatisfied uniqueness-presupposition as in (25) were read slower than test sentences in (24). Thus, when the uniqueness-presupposition of the definite determiner is not satisfied people experience processing difficulties quite early, that is before the end of the sentence. However, the experiment showed that participants did not determine the unsatisfied presupposition immediately on the trigger itself. It seems as if the processing of the presupposition is delayed until later. This could be an artefact of the experimental design because during the experiment a strategy of a delayed presupposition verification turned out to be useful since it was always the case that the presupposition verification was possible at the region following the noun phrase. If this was not the case earlier effects already on the trigger itself could be expected. Nevertheless, the experiment revealed that unsatisfied presuppositions cause processing difficulties.

Van Berkum, Brown, Hagoort, and Zwitserlood (1999) also investigated referentially ambiguous noun phrases and observed early effects when the uniqueness-presupposition was not met. Later on, they replicated their findings with an EEG study and did not observe a delayed verification of the presupposition as Altmann and Steedman (1988) did.

Tiemann et al.'s (2011) second experiment focused on the verification process of presuppositions. Their design made it possible to compare the same test sentence under a verified and a falsified presupposition and found increased reading times in case of unsatisfied presuppositions. Two different context sentences (see (26) and (27)) were paired each with two test sentences (a and b) in such a way that the presupposition of the test sentence was verified (see (26a) and (27a)) or falsified (see (26b) and (27b)) by the corresponding context. In case the presupposition was verified by the first context it was falsified by the second context or vice versa.

(26) Context: Susanne hat dieses Jahr bereits rote Handschuhe gekauft.

Susanne has already bought red gloves this year.

a. Heute hat Susanne wieder rote Handschuhe <u>gekauft</u> und sie gleich angezogen.

Today, Susanne bought red gloves again and put them on right away.

b. Heute hat Susanne wieder keine roten Handschuhe <u>gekauft</u> und ärgert sich.

Today, Susanne didn't buy red gloves again and is very upset.

(27) Context: Susanne hat bisher nie rote Handschuhe gekauft.

Susanne had never bought red gloves until now.

a. Heute hat Susanne wieder keine roten Handschuhe <u>gekauft</u> und ärgert sich.

Today, Susanne didn't buy red gloves again and is very upset.

b. Heute hat Susanne wieder rote Handschuhe <u>gekauft</u> und sie gleich angezogen.

Today, Susanne bought red gloves again and put them on right away.

The data revealed higher ratings in the verifying than in the falsifying condition. The reading times on the critical word (the region where the presupposition is known) were significantly longer in the falsifying compared to the verifying condition. This indicates that a verification process is started as soon as the content of the presupposition is known and that it takes longer in case this process does not succeed.

Further evidence for presuppositions triggered by *auch* (Engl.: 'too') comes from Schwarz (2007). He observed that reading times were slower in contexts that did not satisfy the presupposition compared to contexts that satisfied it.

3.3.2 EEG

Van Berkum, Brown, and Hagoort (2003) replicated previous findings by van Berkum et al. (1999) with two event related potential (ERP) studies and auditory stimuli. The data revealed that definite noun phrases evoked early ERP effects when the uniqueness-presupposition was not met. In both experiments, they presented discourses that either introduced a unique referent (one salient girl as in (28a)) or multiple referents (two equally salient girls as in (29a)). In the test sentence, a particular noun phrase (e.g., the girl) either uniquely referred to a single referent mentioned in the discourse before (see (28b)) or to one of the two introduced referents (see (29b)). Only after the disambiguating relative clause („…that had been on the phone… ") it becomes clear which referent is meant.

(28) a. David had told the boy and the girl to clean up their room before lunchtime. But the boy had stayed in bed all morning and the girl had been on the phone all the time.

b. David told **the girl** that had been on the phone to hang up.

(29) a. David had told the two girls to clean up their room before lunchtime. But one of the girls had stayed in bed all morning and the other girl had been on the phone all the time.

b. David told **the girl** that had been on the phone to hang up.

In both experiments, the definite noun phrase evoked early ERP effects when the uniqueness-presupposition was not met. Those effects arose already on the noun *girl*. Therefore, van Berkum et al. (2003) suggested that referential ambiguity is detected very early during sentence processing. From a pragmatic point of view, this can also be interpreted as unsatisfied presuppositions of definite determiners that lead to early effects in ERP. In contrast the data reported by Altmann and Steedman (1988), the processing of the presupposition was not delayed, but instead participants verified the presupposition in the context as early as they heard/read the noun, although they knew that there were also sentences where the unsatisfied presupposition was resolved by the following discourse. The ERP data suggests that the presupposition of the definite determiner (*the*) is checked as soon as possible and does not depend on the following discourse. Furthermore, the data shows that people realize the presupposition failure of a definite noun phrase immediately on the noun phrase itself and that this process is more difficult than the verification of the presupposition.

Additional evidence from ERP studies comes from Burkhardt (2006). A test sentence with a definite noun phrase (like in (31)) was presented in different contexts that either explicitly introduced the mentioned individual (*given* condition, see (30a)), made it easy to infer the mentioned individual (*bridged* condition, see (30b)) or made it impossible to infer the mentioned individual (*new* condition, see (30c)). The given condition explicitly satisfies the existence presupposition of the definite determiner, while this is not the case for the other two conditions (although it can be easily inferred in the bridged condition).

(30) a. Tobias visited a conductor in Berlin.

b. Tobias visited a concert in Berlin.

c. Tobias talked to Nina.

(31) He said that the conductor was very impressive.

The ERP data revealed an N400 in the new condition, while this effect was less pronounced in the bridged condition. The results are parallel to van Berkum et al. (2003) and support the idea that when the existence presupposition of the definite determiner is not given, an N400 emerges. In cases where the relevant individual can be easily inferred, the effect is less strong.

Furthermore, Burkhardt (2006) found a late positive effect (P600) in the new and in the bridged condition, but not in the given condition. This can be interpreted as a full integration of the respective discourse units at this point. The independent discourse referent is already identified to be stored and maintained in discourse representation. From a pragmatic point of view, this could be regarded as presupposition **accommodation.**

Results observed by Kirsten et al. (2014) revealed further evidence for additional processing costs in case of infelicitously used determiners in a word-by-word reading experiment. Their EEG data revealed that participants recognized the mismatching condition already when reading the determiner. The infelicitous use of the determiners resulted in an N400 which is interpreted as an index for a semantic mismatch, followed by a P600 (after noun onset). The authors interpreted the P600 as the evaluation of possible alternative sentence interpretations.

Hertrich et al. (2015) used similar text material as Kirsten et al. (2014), but presented the stimuli auditorily via headphones and conducted cross-correlated analysis of magnetoencephalography (MEG) signals. Furthermore, Hertrich et al. (2015) also investigated the existence presupposition of the definite determiner and the novelty assumption of the indefinite determiner. They additionally constructed contexts that introduced a single item (e.g., “somebody has a house”) or a non-existing item (e.g., “somebody has no house”) and combined it with test sentences that either referred to an existing item (existence presupposition of the definite determiner as in “Somebody painted the house”) or introduced a new item with an existence-creating verb (novelty presupposition of the indefinite determiner as in “Somebody built a house”). The combination of the definite determiner in contexts introducing a single item or the indefinite determiner in the non-existing contexts resulted in matching conditions. By contrast the combination of the definite determiner in the non-existing condition resulted in mismatching condition (same for the indefinite determiner in the existing context). The acceptability rating yielded low acceptability ratings in the mismatching conditions. The MEG data revealed context-matching effects as early as 50ms after syllable onset following the presupposition trigger. In the mismatching condition the latencies for the M50c deflection were larger compared to the matching condition. The authors interpreted this observation as an indicator for an inhibition of automized phonological/lexical processes in cases where discourse coherence is violated. Furthermore, there was a bi-phasic time course of alpha suppression in mismatching conditions visible which might reflect an immediate processing of the presupposition trigger and a subsequent semantic re-interpretation due to the discourse incoherence. This re-interpretation does not take place in the matching conditions and thus the infelicitous use of determiners comes with additional processing costs (see also Jouravlev et al., 2016, for evidence for the trigger *again*).

3.4 Effect 3: Indefinite determiners are more difficult in processing than definite determiners

As pointed out in Chapter 2, there are several possible explanations, how the anti-uniqueness of the indefinite determiner arises. The experimental literature provides evidence for the anti-presupposition account which suggests that while a sentence with a definite determiner presupposes uniqueness directly, the anti-uniqueness of the indefinite determiner arises only indirectly. According to the Maximize Presupposition principle (Heim 1991) anti-uniqueness results in a two steps: (i) the evaluation of the uniqueness-presupposition of the definite determiner and (ii) the subsequent negation of exactly this presupposition. This means that encountering an indefinite determiner is associated with additional processing costs compared to the evaluation of the definite determiner through the additional step of negating the uniqueness-presupposition.

3.4.1 Reading times

One of the first studies investigating the processing of references was run by Murphy (1984), in which participants read stories sentence by sentence. In his first experiment, he focused on the novelty aspect of indefinite determiners and compared noun phrases in contexts that only differed with regard to the used determiner (definite vs. indefinite determiner; e.g., "Later, George was passed by a/the truck, too"). The reading times of the indefinite determiner were significantly longer than those of the definite determiner, supporting the idea that finding an antecedent for a definite reference is easier than establishing a new referent in the discourse model. The effect was replicated in his second experiment that used pronouns (e.g., "Sandy rode a bicycle and I rode it, too." vs. "Sandy rode a bicycle and I rode one, too.") and contrasted the indefinite pronoun "one" with the appropriate definite pronoun.

Further evidence was provided by Clifton (2013) who reported in two of four experiments more processing difficulties for indefinite than for definite determiners. Participants were presented with context sentences that established whether there are one or multiple of the relevant items (e.g., "In the kitchen …" vs. "In the appliance store…") and the following test sentence contained either a definite or an indefinite determiner ("The stove…" vs. "A stove…"). Reading times in contexts that stereotypically provided a single possible referent for the definite determiner phrase or multiple possible referents for an indefinite determiner phrase were shorter than when context and type of determiner did not match (e.g., definite determiner in a context providing multiple items). It is noteworthy that these results were only obtained when participants were required to perform an additional simple arithmetic task in between reading the sentence and answering a question about it (Exp. 2 and 4). By this additional task, Clifton forced the participants to comprehend the sentence to a deeper level, because he prevented rehearsing the sentence until the question was presented.

3.4.2 Eye-tracking

The idea of an anti-uniqueness inference of the indefinite determiner arising in a two-step process based on Maximize Presupposition is supported by an eye-tracking experiment of Bade and Schwarz (2019b). These authors observed differences between both determiners in a picture selection task. Participants heard test sentences like “A/the shirt in Benjamin’s closet is blue” and saw three different pictures of a boy with a closet. Participants were asked to choose the picture they thought was corresponding to the sentence, and indeed they looked at the respective target picture (picture with a single shirt for the definite determiner, and picture with multiple shirts for the indefinite determiner) very early on upon hearing the noun. Inferences based on the use of the indefinite determiner were drawn to a much lesser degree than those evoked by the use of the definite determiner, as demonstrated by fewer target choices for the indefinite than for the definite determiner. Further differences between the determiners were observed in eye-tracking patterns for the cases where the target was chosen.

3.4.3 EEG

Kirsten et al. (2014) additionally revealed general differences between the definite and the indefinite determiner, namely that the definite determiner evoked an enhanced negativity compared to the indefinite determiner. Test sentences either used a definite or an indefinite determiner and both were combined with context sentences that either satisfied the uniqueness-presupposition of the definite determiner or left it unsatisfied. The data suggest that the underlying processes between the two determiners differ. Similar results were obtained in an MEG study with auditorily presented test sentences (Hertrich et al., 2015).

3.5 Effect 4: Presupposition accommodation is demanding

The accommodation process of presuppositions starts as soon as the content of the presupposition is known and is more demanding than the verification of presuppositional content.

3.5.1 Reading times

Haviland and Clark (1974) provided evidence for processing costs of accommodation. They investigated the comprehension of sentences introducing a new item in comparison to already known information. In a reading experiment participants were instructed to read sentences in different contexts. The target sentence always contained a definite noun phrase. In one condition the context explicitly mentioned the existence of the noun phrase (see (32a)), while this was not the case in the second condition (see (33a)).

(32) Direct Antecedent

a. Context: We got some beer out of the trunk.

b. Target sentence: The beer was warm.

(33) Indirect Antecedent

a. Context: We checked the picnic supplies.

b. Target sentence: The beer was warm.

Participants were instructed to press a button as soon as they understood what the target sentence means. As predicted, the comprehension time of the target sentences in the direct antecedent condition was faster than the comprehension time in the indirect antecedent condition, because in the direct antecedent condition there is a direct referent available in the memory. This is not the case in the indirect antecedent condition. The listener searches the memory for a suitable referent, but this process fails and the new information has to be integrated into the memory. To comprehend the sentence, it is necessary to accommodate the new information which takes additional time. With a second experiment, they excluded the possibility of a facilitating effect of repetition and observed the same results. In a third experiment, they replicated the results with other presupposition triggers including *still*, *either*, *again,* and *too*.

Domaneschi and Di Paola (2018) also investigated processing of presupposition accommodation in a self-paced reading experiment. While one context satisfied the presupposition, another one was neutral towards the presupposition. Besides definite descriptions, they included change of state verbs, iterative expressions and focal particles in the experiment (conducted in Italian). They observed longer self-paced reading times already on the presupposition trigger in contexts where the presupposition is not yet known. Furthermore, they concluded that accommodation is more difficult than satisfaction across triggers.

3.6 Summary

In the present chapter, I introduced important experimental methods that are used to investigate presupposition processing. So far, acceptability ratings, reading time experiments, eye-tracking studies, and EEG experiments were used to investigate presupposition processing. For this dissertation, four experimental results are of special interest: First of all, the processing of presuppositions starts immediately, which means already on the trigger. This leads to additional processing costs for presupposition triggers compared to neutral words. Second, failure of the verification process of the presupposition content leads to additional processing cost and thus infelicitously used presupposition triggers evoke additional processing difficulties compared to felicitously used ones. Third, processing indefinite determiners is more difficult than processing definite determiners, because the anti-uniqueness inference of the indefinite determiner arises in two steps which involve considering the uniqueness-presupposition of the definite determiner in its first step and its subsequent negation in a second step. Finally, the accommodation of the presuppositional content starts as soon as it is known and this process is more difficult than the verification of a presupposition.

In the following chapter, I will introduce a paradigm from cognitive psychology: The Psychological Refractory Period approach. The combination of this paradigm and self-paced reading experiments made it possible to investigate the nature of cognitive processes involved in presupposition processing.

4. Psychological Refractory Period Paradigm

The Psychological Refractory Period (PRP) approach has been widely used in cognitive psychology and has its origin in dual-task research (Pashler, 1984, 1994; Telford, 1931). In combination with the locus of slack-logic (going back to Schweickert, 1978) it is an accepted way to determine whether a process is automatic or requires limited capacities.

In a PRP experiment, participants perform two independent tasks in each trial. The two tasks require separate responses (R1 and R2) that have to be given in a predetermined order. The critical manipulation is the stimulus onset asynchrony (SOA), the time between the two stimuli (S1 and S2). With a short SOA the two tasks temporally overlap, whereas with a long SOA there is no or only little temporal overlap. Typically, SOA has no (or only little) influence on the response time in Task 1 (RT1), but the response times in Task 2 (RT2) increase, when SOA decreases, thus reflecting dual-task interference. This is the so-called PRP effect (Telford, 1931) which can be quantified as the difference between RT2 at the long and at the short SOA. Welford (1952) and Pashler (1994) explained this effect with the central bottleneck model (CBM) (see Figure 1 for an illustration). According to this model, a task is divided into three stages: i) a pre-central stage, ii) a central stage, and iii) a post-central stage (see Figure 1 (a)).

The pre-central stage has often been related to perceptual processing and the post-central stage to motor processing and execution. The core assumption of this model is that while the pre- and the post-central stage can be processed in parallel with other stages, only one central stage can be processed at any time because it is conceived capacity-limited and thus it constitutes a bottleneck. As a consequence, with a short SOA the central stage of Task 2 has to be postponed until the central stage of Task 1 is processed. The resulting waiting time is called *cognitive slack* and leads to an increase in the corresponding RT2. In case of a long SOA, there is no (or only little) cognitive slack, because the central stage of Task 1 is already processed and thus processing of Task 2 is not interrupted, resulting in shorter RT2s (see Figure 1 (a)).

The PRP paradigm can be used to investigate at which stage of processing a certain RT effect emerges. One way to do so is using the locus of slack-logic (Schweickert, 1978). It allows to distinguish whether a particular process that is responsible for an RT effect can run in parallel before the bottleneck (in the pre-central stage) or whether this process runs in the central or post-central stage. In case the effect arises during the central stage and thus requires limited-capacities, the effect becomes visible in RT2 at both long and short SOAs to the same degree.[1] Consequently, the manipulation (that affects the central stage of the second task) and SOA should combine additively (see Figure 1 (b)). In case the effect results from parallel processing during the pre-central stage, it can run in parallel to the central stage of Task 1 and thus it can already be processed even with a short SOA. Thus, the manipulation affects the pre-central stage and its effect on RT2 varies with SOA. With a sufficiently short SOA, the effect is absorbed in the cognitive slack and the expected difference between conditions in Task 2 cannot

[1]Strictly speaking, the same additive pattern is observed when the effect arises from the post-central motor stage. However, identifying whether the locus of the effect is in the central or post-central stage of processing is impossible with the introduced setting. In case it can be ruled out that the effect has its locus in the pre-central stage, the effect propagation-logic can be applied (e.g., Durst & Janczyk, 2018; Janczyk, Humphreys, & Sui, 2019; Miller & Reynolds, 2003). This logic allows to differentiate between the locus originating in the post-central stage or in an earlier stage. Combining those result makes it possible to identify the locus of the effect under investigate.

be observed any longer. In case of a long SOA, the central stage of Task 1 is already processed and there is no (or only little) cognitive slack. The processing of the effect prolongs RT2s and the RT effect between conditions in Task 2 becomes visible (see Figure 1(c)). Consequently, the manipulation and SOA interact underadditively.

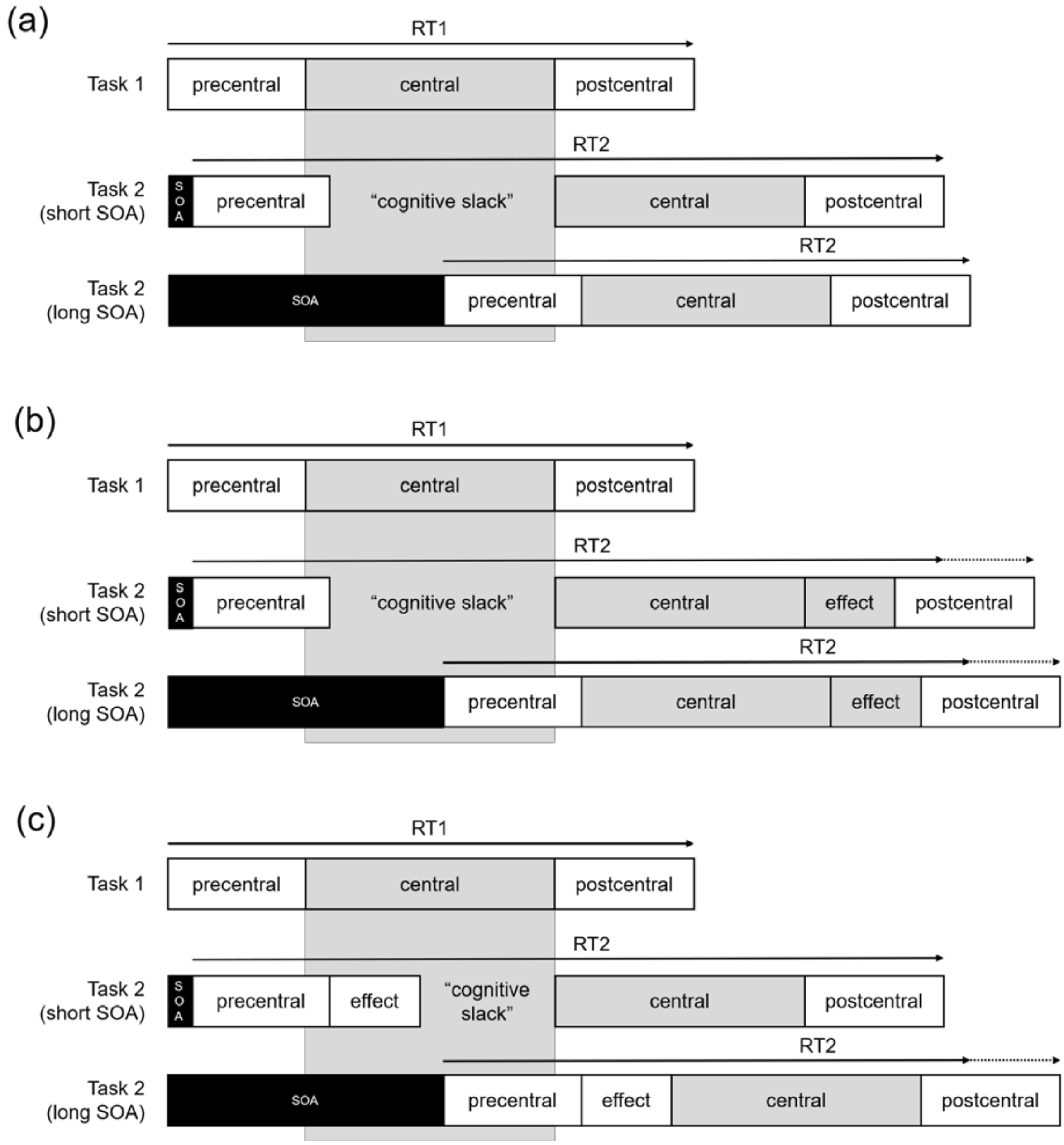


Figure 1: Illustration of the CBM (a) and the predictions of the locus of slack-logic (b) and (c).

There also other approaches to explaining the PRP effect with other models than the CBM. For example, Navon and Miller (2002) and Tombu and Jolicoeur (2003) have suggested capacity sharing models that allow for parallel processing of the central stage. Their approach also assumes a capacity-limitation: While parallel processing is possible, the available capacity must be divided between tasks leading to performance decrements in turn. Tombu and Jolicoeur (2003) determine the exact proportion by a sharing parameter in their model. When all capacity is first devoted to Task 1 processing, their model mimics the CBM. Nevertheless, predictions for Task 2 are similar to those made by CBM, even in less extreme scenarios. To reduce complexity, I derive and illustrate predictions only from the previously introduced CBM. The predictions that can be derived for the first task are indeed different depending on the applied model, but the main focus of the conducted experiments is on the performance of Task 2.

These methods have been used in many studies on different topics, for example the origin of the Stroop effect (Stroop, 1935). The Stroop effect means longer RTs in color naming the ink color in incongruent conditions (e.g., *blue* printed in red ink), compared to congruent condition (e.g., *red* printed in red ink). For instance, Fagot and Pashler (1992) concluded that the Stroop effect arises during the central stage of response selection and thus requires limited capacities. Piai, Roelofs, and Schriefers (2014) applied the locus of slack-logic to investigate picture-word interference (PWI) where participants are instructed to name the pictures and to ignore the distractor that can have a, for example, semantic or phonological relation to the word. A series of experiments revealed additive effects of SOA and stimulus type on picture naming RTs. The authors argued according to the observed data against a locus at the pre-central stage of the semantic interference (e.g., Ayora et al., 2011; Dell'Acqua, Job, Peressotti, & Pascali, 2007).

The automaticity of the learning process during study of implicit learning (artificial grammar learning; AGL) was investigated by Hendricks, Conway, and Kellogg (2013). In a dual-task setting they investigated the contribution of automatic and capacity-limited processes involved in AGL. According to the data, some aspects of learning in AGL are relatively automatic and therefore can run in parallel to the central stage of another task, but the expression of grammatical information and the learning of grammatical patterns appear to require central capacities.

5. Research questions

In Chapter 3, I presented relevant literature concerning presupposition processing. In Chapter 4, the PRP method and its use in cognitive psychology was introduced. In the present chapter, I will motivate the three questions that were investigated the current dissertation.

First of all, I compared processing of definite and indefinite determiners and investigated how the anti-uniqueness inference of the indefinite determiner arises. Secondly, I applied the PRP paradigm to delineate the nature cognitive processes required for presupposition processing, especially for presuppositions triggered by determiners. Due to the interdisciplinarity of the present research project, I decided to take an unconventional path and combined the PRP paradigm with self-paced reading. This novel approach made it possible to investigate the required capacities and it sheds light on the underlying processes involved in presupposition processing. Finally, I validated previous results with larger sample sizes and focused on the question *when* presupposition processing, especially triggered by determiners, takes place. With five experimental studies, I provided data that allow to preliminary answer these questions.

The **main question** of the present dissertation was: What are the underlying processes of presupposition processing? This question can be broken down into three sub-questions:

Q1: How does the anti-uniqueness interference of the indefinite determiner arise?

Q2: What cognitive resources are involved during presupposition processing?

Q3: When do presuppositions unfold their impact?

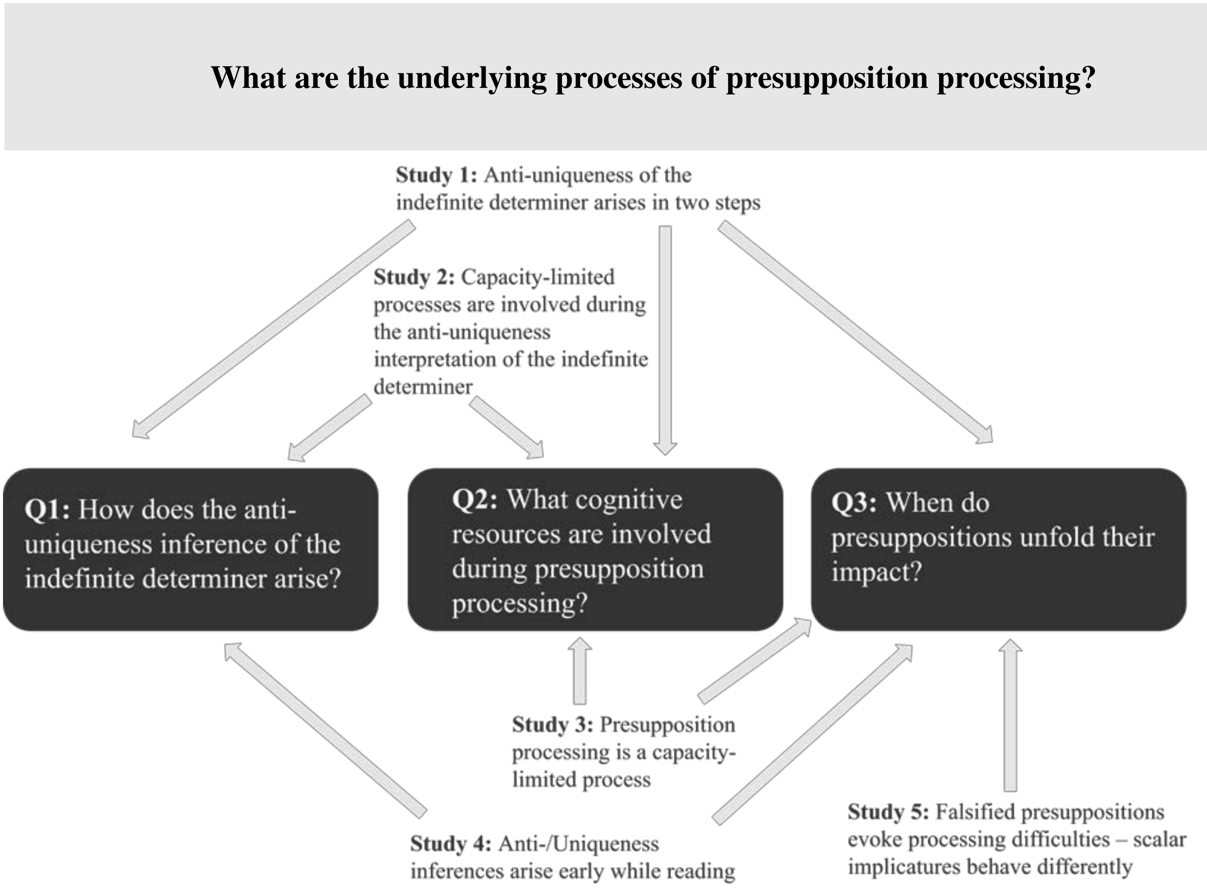


Figure 2: Illustration of the research questions and studies that provide evidence for each question.

Figure 2 visualizes the interplay of the five studies and the sub-questions addressed in this dissertation. In the following chapter, I will summarize the five studies and their results.

6. Studies

6.1 Study 1: Anti-uniqueness of the indefinite determiner arises in two steps[2]

In this study, I conducted two mouse-tracking experiments to compare processing of definite and indefinite determiners and to gather evidence for the anti-presupposition theory based on the Maximize Presupposition principle.

Before going into detail about the present experiment, I will introduce the method of mouse-tracking in general. Hand movements have been shown to be very sensitive to reveal the temporal dynamics of cognitive states (Song & Nakayama, 2009; see also Freeman, Dale, & Farmer, 2011). Recording movements of the computer mouse on its way to a particular goal-location representing a certain decision is an easy way to gather the required continuous data, and has become increasingly popular in cognitive psychology and experimental linguistics. In a standard mouse-tracking task, participants are presented with an image, letter, string, sound, video (or a combination of these) and then make a response via moving the cursor to options that (usually) appear in the top left or top right corner of the computer screen. While participants move the mouse cursor, the x- and y-coordinates of the cursor en route to the response are recorded. Typically analyzed measures are the area under the curve (AUC) (Freeman & Ambady, 2010), which is the area between the observed trajectory and an idealized straight line from the start to the end point. X_{neg} is the amount of horizontal deviation toward the competitor (i.e., toward the non-target response) and is measured as the maximum x-coordinate reached by

[2]Based upon: Schneider, C., Schonard, C., Franke, M., Jäger, G., & Janczyk, M. (2019). Pragmatic processing: An investigation of the (anti-)presupposition of determiners using mouse-tracking. *Cognition, 193*, 104024

the mouse as it veers away from the target response (Tomlinson, Bailey, & Bott, 2013). As temporal measures one can analyze the movement time which is the time from stimulus onset until the cursor reaches the response box (Kieslich & Henninger, 2017) and turn towards target (TTT) which is the time when participants finally make their decision and turn towards the target response without any subsequent reversals (Roettger & Franke, 2019). Figure 3 illustrates the analyzed dependent variables.

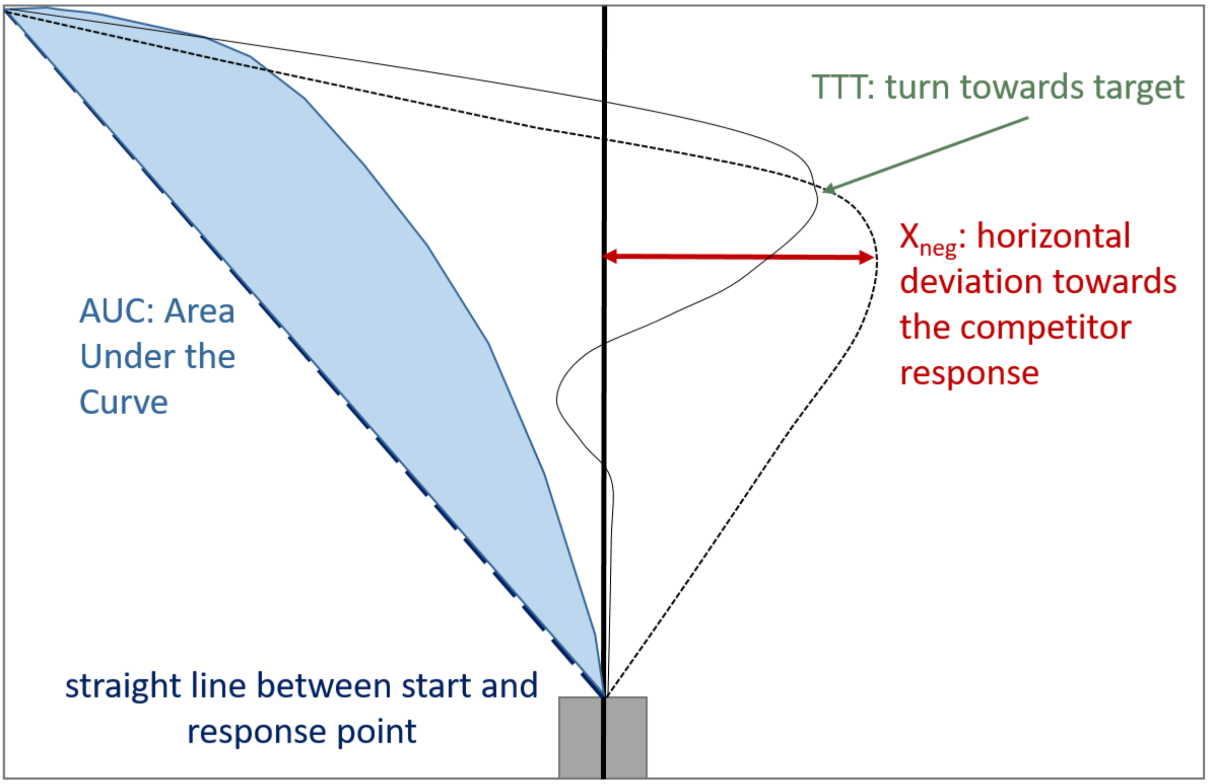


Figure 3: Illustration of the dependent variables AUC, X_{neg}, and TTT

The aim of Study 1 was to analyze the unfolding of processing of the definite and the indefinite determiner during a mouse-tracking experiment. In both experiments participants were instructed to judge the appropriateness of test sentences in a given context. The context consisted of a box (the "shopping basket") with three pieces of fruit additionally described by a sentence (e.g., "Jan's mum was shopping. She bought one banana and two pears."). Furthermore, there was a picture of Jan on the screen.

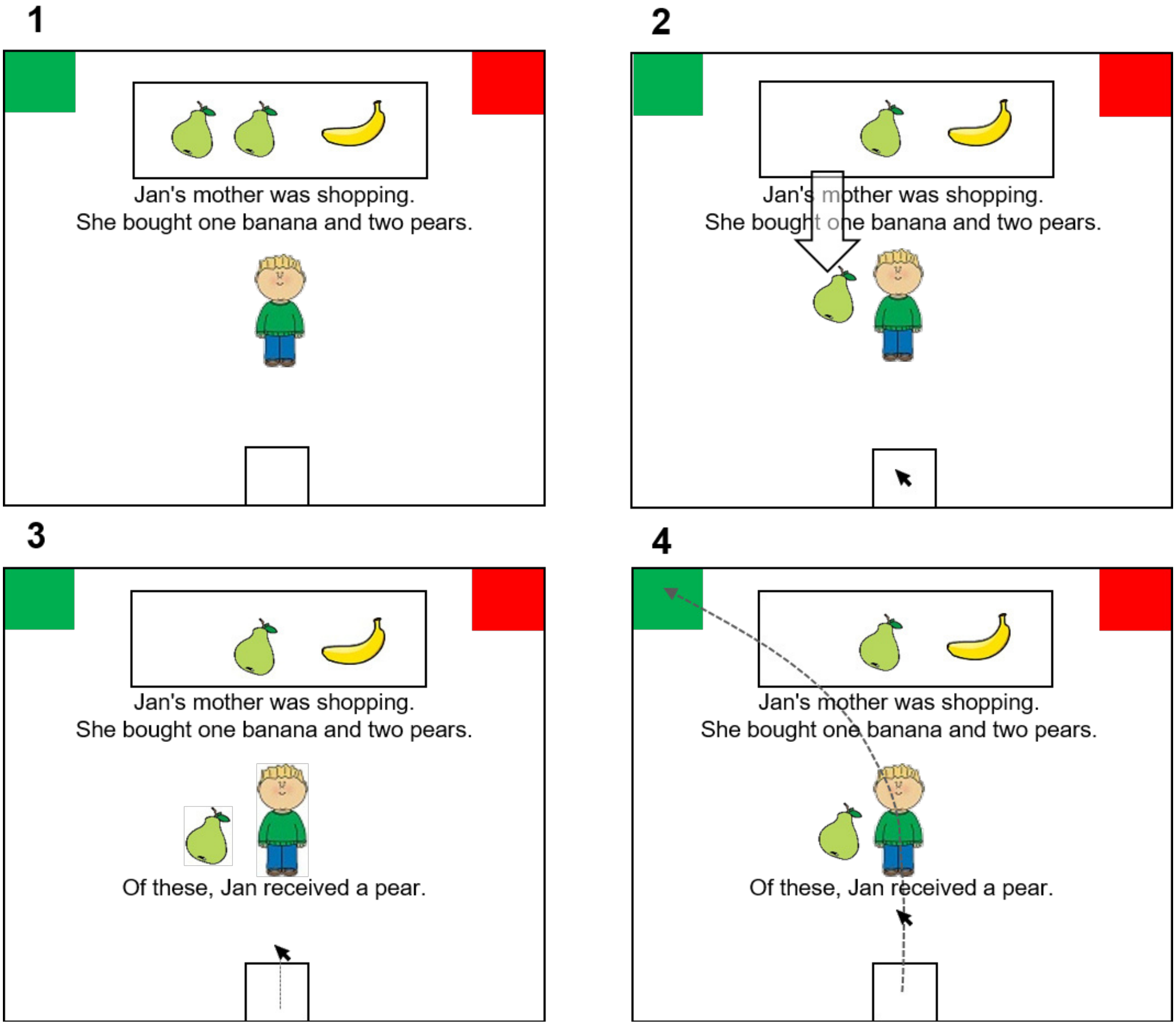


Figure 4: Illustration of the task in Study 1.

To start a trial, participants had to move the mouse cursor into the start box. Then, Jan received one piece of fruit from the shopping basket. As soon as the mouse cursor was moved into the upper direction the test sentence appeared on the screen (e.g., "Of these Jan received a pear."). Participants had to judge the test sentence according to the presented context as "true" or “false” by moving the mouse in the corresponding response box into the upper left or upper right corner of the screen (for an illustration of the procedure, see Figure 4).

Six conditions were used in the experiment and they resulted from crossing the two determiners (definite and indefinite) with three sentence types (false, felicitous, and infelicitous) (for more information see Figure 5). The false conditions were integrated to create a baseline.

Context: Jan’s mother was shopping. She bought one banana and two pears. 

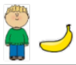
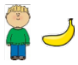
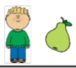
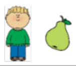
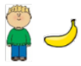
Sentence type	Determiner	Picture	Stimulus sentence	Expected response
false	definite		Of these, Jan received **the** banana.	false
	indefinite		Of these, Jan received **a** pear.	false
felicitous	definite		Of these, Jan received **the** banana.	true
	indefinite		Of these, Jan received **a** pear.	true
infelicitous	definite		Of these, Jan received **the** pear.	false
	indefinite		Of these, Jan received **a** banana.	false

Figure 5: Examples for the six different conditions of Study 1.

Following the anti-presupposition theory, the anti-uniqueness interpretation of the indefinite determiner is derived indirectly in a two-step process, by first assessing the uniqueness-presupposition of the definite determiner and then negating it in a second step. For the conditions in which the presupposition matters (condition 3-6) this leads to the predictions of larger AUC and longer MT in the indefinite compared to the definite conditions. Most importantly, the two-step model of processing of the indefinite determiner requires an initial consideration of the uniqueness-presupposition of the definite determiner and its negation in a second step thus predicting larger X_{neg} values for the indefinite relative to the definite determiner. To explain this prediction in more detail, consider first the definite determiner. In the felicitous condition the uniqueness-presupposition of the definite determiner is supported by the context and the mouse-cursor directly heads towards the expected “true” response. In the infelicitous condition the uniqueness-presupposition of the definite determiner is not supported by the context and thus the mouse movement directly heads towards the expected “false” response. For the indefinite determiner a different pattern is expected. In the felicitous condition the uniqueness-presupposition of the definite determiner is expected to be initially activated, but it is not supported by the context and thus a bias towards the “false” response is expected before the expected “true” response is selected. In the infelicitous condition, uniqueness is supported by the context and due to the initial activation of the uniqueness-presupposition a bias towards a “true” response is expected, even when finally, the “false” response is given. This initial bias towards the unselected response leads to higher X_{neg} values for the indefinite compared to the definite determiner.

In the first experiment, participants were provided with error feedback throughout the whole experiment, while in the second experiment error feedback was only provided during a practice phase. Across both experiments, the indefinite determiner was associated with more processing difficulties which was reflected in longer MTs, larger AUC, and larger X_{neg} values for the indefinite determiner. The larger X_{neg} values support the evaluation of the anti-uniqueness of the indefinite determiner according to the Maximize Presupposition principle (Heim, 1991). According to this principle, the anti-uniqueness of the indefinite determiner is an anti-presupposition and thus arises in two steps: first the uniqueness-presupposition of the definite determiner is considered, which is negated in a second step and thus leads to anti-uniqueness. This initial consideration of the uniqueness-presupposition leads to a bias to the nontarget response in case of the indefinite determiner and is thus reflected in larger X_{neg} values in felicitous and infelicitous conditions. It is important to point out that the initial bias towards the nontarget was also observable in the infelicitous condition although the context facilitates a “false” response. One has to point out that this initial consideration of the uniqueness-presupposition is an unconscious process during sentence processing.

Both experiments provide experimental evidence for the psychological reality of the Maximize Presupposition principle and the assumed two-step processing of the anti-uniqueness inference of the indefinite determiner. Processing difficulties of infelicitous conditions were observed in Experiment 1, but were only replicated for MTs in Experiment 2. The included false conditions might have influenced this result and thus the unclear findings were further investigated in Study 2.

6.2 Study 2: Capacity-limited processes are involved during the anti-uniqueness interpretation of the indefinite determiner[3]

The second study aimed at replicating the increased processing difficulties for the indefinite determiner and to clarify the inconsistent results observed for the infelicitous conditions in the previous study. Furthermore, I expanded the research question and asked whether processes involved during the interpretation of the indefinite determiner can either run in parallel with the central stages of another task or whether they are capacity-limited and require the central bottleneck itself. This is of interest because when treating the anti-uniqueness of the indefinite determiner as an anti-presupposition, it is expected that processing of the indefinite determiner takes longer because it involves the evaluation of the uniqueness-presupposition of the definite determiner and its subsequent negation. Previous research has shown that the process of linguistic negation is a "resource dependent process (Strack & Deutsch, 2004, p. 227; see, e.g., Deutsch, Kordts-Freudinger, Gawronski, & Strack, 2009, Exp. 3; Foerster, Wirth, Berghoefer, Kunde, & Pfister, 2019; Wason, 1959). In case negation is involved in the process of interpreting the indefinite determiner, this process cannot run in

[3] Based on Schneider, C., & Janczyk, M. (accepted). Capacity limitations of processing presuppositions triggered by determiners. *Acta Psychologica.*

parallel to other central stages. To investigate this question, I used the same test sentences as in Study 1 except for the false conditions, and employed the Psychological Refractory Period (PRP) approach and the locus of slack-logic. As Task 1, participants were instructed to perform a tone discrimination task before reading the determiner. The sentence was completed after a short SOA (100ms) or a long SOA (1200ms) and as Task 2 participants had to evaluate the appropriateness of the sentence against the presented context. Both tasks were performed with button presses with the left and right index and middle finger (see Figure 6 for an illustration of the task).

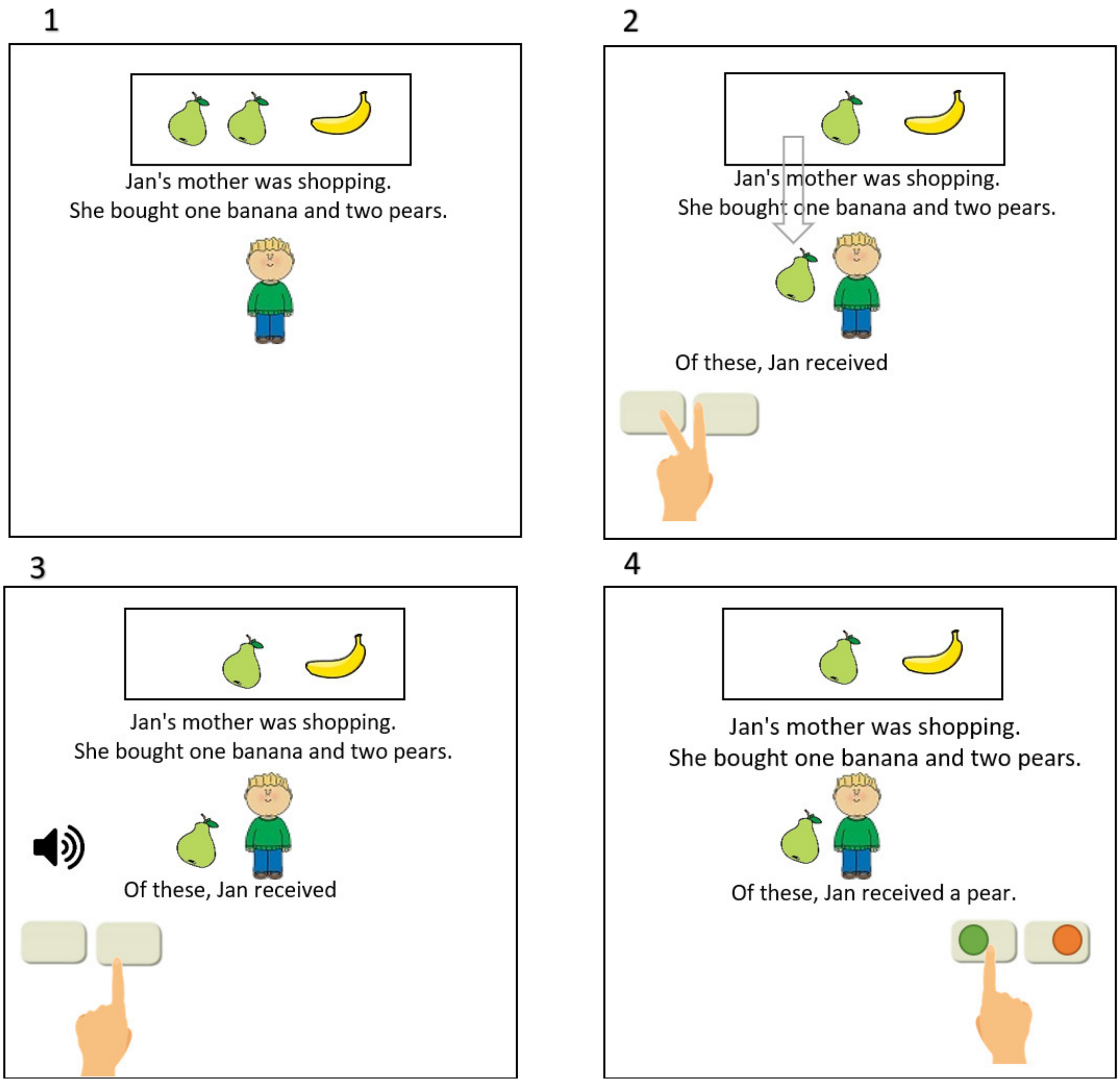


Figure 6: Illustration of an example trial in in Study 2.

In case the processing of the indefinite determiner requires an additional capacity-limited process (compared to the definite determiner), that is negation, I expected to observe a difference between the definite and the indefinite determiner with a short and a long SOA (i.e., an additive combination of SOA and the determiner variable). In case the processing does not require limited capacities and can run in parallel to the central stage of the tone discrimination task, I expected to observe a difference between the two determiners only in case of a long SOA, but not with a short SOA (i.e., an underadditive interaction). With the short SOA the effect is absorbed in the cognitive slack and thus not visible in RT2. Such a result would be not in line with the Maximize Presupposition principle.

The data clearly revealed two main effects of determiner and sentence type which is well in line with the hypotheses. Processing of the indefinite determiner took longer than processing of the definite determiner and responding in felicitous conditions was faster than in infelicitous ones. Both effects combined additively with the SOA manipulation which fits well with the anti-presupposition account based on the Maximize Presupposition principle (Heim, 1991). The anti-uniqueness inference of the indefinite determiner arises in two steps, namely the evaluation of the uniqueness-presupposition of the definite determiner and its subsequent negation. In terms of the locus of slack-logic one or both of these processes require the central bottleneck and are thus capacity-limited. As negation is regarded as a capacity-limited process (e.g. Deutsch et al., 2009; Foerster et al., 2019; Strack & Deutsch, 2004; Wason, 1959) it is a viable candidate and thus the results provide experimental evidence for the Maximize Presupposition principle.

However, determiner and sentence type interacted: In infelicitous sentences the RTs for definite and indefinite determiners were almost the same. This might be the result of a speed-accuracy tradeoff (e.g. Liesefeld & Janczyk, 2019), because the error rates were larger for the indefinite compared to the definite determiner across all conditions. Furthermore, the infelicity of the definite determiner might be perceived larger than for the indefinite determiner, because the definite determiner is even more implausible in a context with multiple objects than the indefinite determiner is in a context with a unique object. This infelicity effect might mask the expected RT difference between the two determiners in the infelicitous condition.

Although the sentence type RT effect was larger for the definite compared to the indefinite determiner, the data revealed that infelicitous uses take longer than felicitous uses and thus this clarifies the somewhat inconsistent results observed in the first study.

In short, Study 2 revealed that during processing the indefinite determiner, a capacity-limited process is involved which is not involved in the evaluation of the definite determiner (namely negation). Study 3 compared the processing of presupposition triggers to other (neutral) words and investigated the question whether presupposition processing itself requires limited cognitive capacities which are not involved during the processing of neutral words.

6.3 Study 3: Presupposition processing is a capacity-limited process[4]

Study 3 investigated determiners, and included a second trigger, namely *again* which is characterized to belong to another class of presupposition triggers than the definite determiner. Experiment 1 was conducted to replicate early effects already on the trigger and to support the need of separate analyses of the different triggers. In a self-paced reading experiment, participants were instructed to read sentences sectionwise and to rate their appropriateness in a given context. I compared sentences including a presupposition trigger (either definite determiner or *again*), with sentences including a neutral word that was grammatically acceptable, but did not trigger a presupposition, and sentences with an unacceptable word making the sentence ungrammatical. The analysis revealed early effects already on the trigger and thus the experiment replicated previous findings by, for example, Tiemann et al. (2011). Nevertheless, the data also showed that separate analyses are necessary because the processing pattern differed between the two triggers. For both triggers reading times at the trigger position were longer for the condition including a presupposition trigger compared to the unacceptable

[4] Based on Schneider, C., Bade, N., & Janczyk, M. (2020). Is immediate processing of presupposition triggers automatic or capacity-limited? A combination of the PRP approach with a self-paced reading task. *Journal of Psycholinguistic Research, 49,* 247-273.

sentence. The neutral sentences of the *again* sentences had the longest reading times whereas for the determiner sentences the neutral sentences were similar to those of the unacceptable condition.

In Experiment 2, I combined the previous experiment with the PRP approach to answer the question of capacity-limited or automatic processing of the two presupposition triggers. I thus included a tone discrimination task before participants read the respective word at the presupposition triggering position. Similar to Study 2, participants were instructed to perform a tone discrimination task after the first part of the sentence (all words preceding the trigger position). The word at the trigger position of the sentence was revealed after a short SOA (100ms) or a long SOA (1200ms). As Task 2, participants had to read the rest of the sentence section wise. Responses to the tone discrimination task were made with the left index or middle finger, the self-paced reading task was performed with the right index finger. This procedure allows to investigate whether presupposition processing requires limited-cognitive capacities that are not involved in the processing of other (neutral) words. If presupposition processing requires limited capacities and thus can only start once the central stage of the (preceding) tone discrimination task has finished, I expected to observe similar differences between the conditions as in the first experiment, which should be of the same size with both the short and the long SOA. In case processing the trigger position can run in parallel to other capacity-limited stages, it extends into the cognitive-slack with a short SOA and thus no difference between the different conditions is observable for the short SOA, but only for the long SOA.

The data analysis revealed a similar pattern of differences between the conditions with the long SOA and thus successfully replicated previous findings. Furthermore, these differences were not absent or smaller with the short SOA[5]. Consequently, the results contradict the notion

[5] In fact, the differences were even larger with a short SOA. This was not predicted by the CBM, but it is discussed in the manuscript in the appendix.

of an automatic processing of presuppositions triggered by definite determiners and *again*. The findings are in line with the assumption that presupposition processing immediately starts a context search for a suitable referent. In case of the definite determiner, an appropriate referent is searched, and in case of *again*, the search for a suitable previous event is immediately started when reading the trigger word. In the current setting, this search process cannot be successful, because the context provided before the test sentence did not explicitly verify the following presupposition. There is no clear referent for the definite determiner introduced in the context, nor is there a previous event mentioned that *again* could refer to. To make sense of the test sentence participants have to accommodate the presupposition, which is likely to be the case as suggested by high ratings in the acceptability rating. This accommodation process is assumed to be a process of enriching the context with the presupposed information (as long as it is contextually feasible, which was the case in the present study). For the preparation of accommodation, cognitive resources like working memory play a role. Previous results by Anderson and Holcomb (2005) also suggested a link between the processing of the uniqueness-presupposition of the definite determiner and working memory.

In sum, the experiment provided evidence for an immediate processing of presuppositions starting already on the trigger. Furthermore, the data suggests that this processing requires limited cognitive capacities in case the presupposition has to be accommodated and is not directly satisfied in a previous context (this topic is part of Study 5).

6.4 Study 4: Anti-/Uniqueness inferences arise early while reading[6]

While Study 2 and 3 focused on the cognitive capacities that are involved during presupposition processing this study addresses whether participants actually use the relevant

[6] Based on Schneider, C., Bade, N., Franke, M., & Janczyk, M. (2020). Presuppositions of determiners are immediately used to disambiguate utterance meaning. *Psychological Research,* https://doi.org/10.1007/s00426-020-01302-7.

meaning components of the definite and indefinite determiner (anti-/uniqueness) to form expectations already while reading it. Furthermore, I wanted to clarify how the anti-uniqueness assumption of the indefinite determiner arises. While the definite determiner clearly conveys a uniqueness-presupposition, the status of the anti-uniqueness inference associated with the indefinite determiner is less clear. As mentioned in Chapter 2, there are, according to the literature, three possibilities: (1) The presupposition theory suggests that the indefinite determiner comes with its own presupposition and thus presupposes anti-uniqueness (see Kratzer, 2005, or the discussion in Heim 1991, 2011). From a processing perspective, no processing differences between the two determiners are then expected. According to the (2) anti-presupposition theory, the anti-uniqueness of the indefinite determiner is derived by considering the (stronger) alternative with the definite determiner and then negating its presupposition. The inferences, which are the result of pragmatic reasoning based on Maximize Presupposition, are not proper presuppositions, but are referred to as anti-presuppositions (Percus, 2006). Because the anti-uniqueness inference is derived in two steps, the processing of the indefinite determiner should thus be more complex than processing the definite determiner. (3) The implicature theory assumes that the indefinite determiner triggers an implicature due to its competition with other quantificational terms, for example, "every/all" or "another" (Chierchia, Fox, & Spector, 2012; Grønn & Sæbø, 2012). Those quantificational terms form a lexical scale with the indefinite determiner (Horn, 1972). An implicature arises when the weaker item on such a scale is chosen, in this case the indefinite determiner. All items that are higher in the scale (items that trigger stronger alternatives) get negated. Following this account, the indefinite determiner should come with an implicature, which has shown to be processed more rapidly (at least if certain conditions are met) than presuppositions (Bill et al., 2018; but see also Chemla, 2008), thus the indefinite determiner should be processed faster than the definite determiner.

The experiment comprised two parts in which a shelf with ten compartments was visible in the upper part of the screen. In the left- and rightmost compartment one or two objects (the same within one compartment, but different objects in the shelf) were visible on the screen. The first part of the experiment was a forced choice production task where participants were asked to produce a sentence which appropriately described a given situation. This task was used to identify participants that were not aware of the uniqueness/anti-uniqueness associated with the two determiners and therefore were replaced with new participants.

The second part was a mouse-tracking experiment with the aim to test whether listeners rapidly integrate potential cues about uniqueness or anti-uniqueness and use this information for disambiguation (even before hearing the lexically disambiguating referent). Furthermore, I compared the processing of definite and indefinite determiners. Similar to Study 1 participants initiated a trial via moving the mouse out of the start box, but in contrast to the previous experiment in Study 1 the test sentence was played via headphones. Participants were instructed to keep on moving the mouse to the left- or right most compartment of the shelf where the corresponding item was displayed (; see Figure 7 for an illustration of a trial).

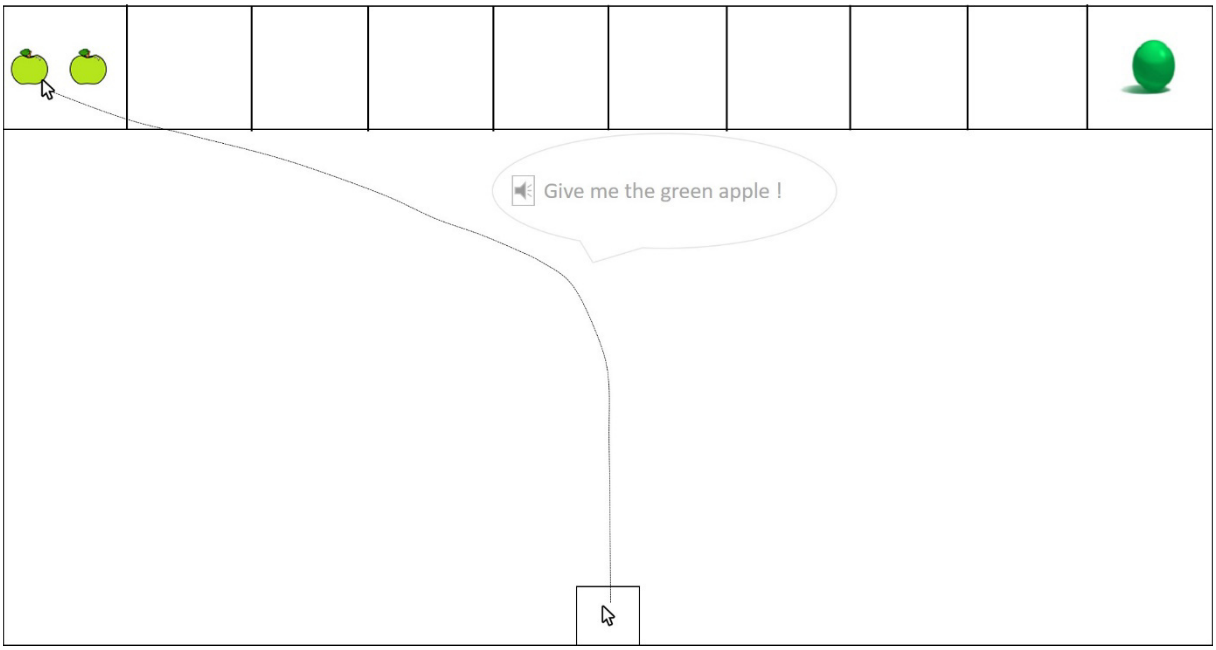


Figure 7: Example of the stimulus setup for a trial in the mouse-tracking task.

The experimental conditions are visualized in Table 1. Participants were divided into two groups: The reliable group only saw items from Conditions 1-4, participants in the unreliable group additionally saw items where the determiners were used infelicitously (Condition 5 and 6). Furthermore, the variable disambiguation indicates whether already the determiner can be used to disambiguate the sentence meaning (early) or whether this is only possible after reading the noun (late). For the mouse-tracking task, I analyzed the dependent measures AUC, MT, and TTT.

Table 1: Summary of the experimental conditions for the mouse-tracking task. The target in this example is always "green apple", and "green umbrella" is the competitor Participants in the reliable group only received trials of Conditions 1-4, the unreliable group received additional trials of Condition 5 and 6.

	determiner	**dis-ambiguation**	**felicity**	**sentence**	**#target**	**#competitor**	**picture**
1	definite	early	felicitous	Give me **the** green apple!	1	2	
2	indefinite	early	felicitous	Give me **a** green apple!	2	1	
3	definite	late	felicitous	Give me **the** green apple!	1	1	
4	indefinite	late	felicitous	Give me **a** green apple!	2	2	
5	definite	early	infelicitous	Give me **the** green apple!	2	1	
6	indefinite	early	infelicitous	Give me **a** green apple!	1	2	

The data revealed in the reliable group smaller AUC values and shorter MT and TTT values for the early disambiguation conditions (Conditions 1-2) compared to the late disambiguation conditions (Conditions 3 - 4). This suggests that participants use the information encoded in the determiner to disambiguate sentence meaning as soon as possible, that is, on the determiner. Further, the number information appears to be encoded in both determiners and is rapidly accessible. However, this was not the case for the unreliable group. It seems as if occasionally infelicitously used determiners made participants stop using the early cues.

The results regarding the origin of the anti-uniqueness inference of the indefinite determiner were less clear though. Overall, the data do not support the implicature theory according to which the processing of the indefinite should be fast compared to the processing of the definite determiner. Regarding AUC and error rates in the production task, there was no significant difference between the two determiners, which supports the presupposition theory where the indefinite determiner comes with its own presupposition of anti-uniqueness. In contrast, for the time base measures MT and TTT, I observed longer times for the indefinite compared to the definite determiner. This supports the anti-presupposition theory according to which the indefinite determiner causes more processing difficulties than the definite determiner due to the additional reasoning processes involved in deriving its inference. In sum, the data support the idea of a rapid use of the inferences in case of felicitously used determiners. Participants are aware of the uniqueness/ anti-uniqueness inference associated with the definite and the indefinite determiner. The origin of the anti-uniqueness inference of the indefinite cannot be clarified conclusively, but the data clearly do not support the implicature theory. The used method in the experiment made it hard to distinguish whether the anti-uniqueness inference of the indefinite determiner results as a presupposition itself, or more indirect as an anti-presupposition. The setup of the experiment made clear that the definite and the indefinite determiner were in competition and that the number of items played a role. Furthermore, participants were directly addressed by the speaker with an order which made clear what the knowledge state of the speaker was. This might increase the possibility of a parallel treatment of the two determiners as predicted by the presupposition account, but it cannot be ruled out that it is due to the setting.

6.5 Study 5: Falsified presuppositions evoke processing difficulties – scalar implicatures behave differently[7]

Study 5 investigated the validation of presuppositions in a given context. Furthermore, scalar implicatures (arising from the Horn scale <some, all>) were included to contrast the processing of two pragmatic phenomena. The aim of the experiment was to replicate previous findings by Tiemann et al. (2011) suggesting an early processing of presuppositions starting already on the trigger. Furthermore, two triggers that are assumed according to the literature to belong to different classes were compared. In addition, the processing of scalar implicatures was investigated and the time-course of processing presuppositions and scalar implicatures was compared.

Two context sentences (A and B) were constructed for each item. Each context sentence was paired with two test sentences in such a way that one context verified the presupposition/scalar implicature of the test sentence and the other context falsified it. Each participant saw each item only in one context condition (but in the verifying and in the falsifying condition). This procedure allows a comparison of exactly the same test sentence in both conditions (verifying and falsifying). In the following there is an example item for definite determiner. The items for *again* were similar to those of determiners and can be found in the manuscript in the appendix. Sentences with scalar *some* (German *einige*) were used for the implicature condition. As a control, sentences with *all* (German *alle*) were included, because they are already most informative no implicature can arise and they can be only literally true or false (an example item can be found in the manuscript in the appendix).

[7] Based on Schneider, C., Janczyk, M., & Bade, N. (submitted). Verifying pragmatic content in context: An experimental comparison of presuppositions of *again* and the definite determiner with scalar implicatures. *Manuscript submitted for publication.*

Example item for definite determiner:

(A) Manuel hat ein Ticket für ein Baseballspiel gekauft.

Manuel bought a ticket for a baseball match.

1. Manuel holt das Ticket und freut sich.

   Manuel collects the ticket and he is happy.

2. Manuel holt die Tickets und freut sich.

   Manuel collects the tickets and he is happy.

(B) Manuel hat mehrere Tickets für ein Baseballspiel gekauft.

Manuel bought several tickets for a baseball match.

3 Manuel holt die Tickets und freut sich.

Manuel collects the tickets and he is happy.

4 Manuel holt das Ticket und freut sich.

Manuel collects the ticket and he is happy.

Participants were instructed to read the sentences in a word-by-word self-paced reading manner and to rate their acceptability according to the given context afterwards. Following previous results by Tiemann et al. (2011), I expected that the falsification of a presupposition causes higher processing difficulties than its verification, thus longer reading times in the falsifying condition compared to the verifying condition are expected. Furthermore, I assumed that the processing of the presupposition triggered by definite determiners and *again* differs because they belong to separate classes. For the trigger *again* a smaller effect of context condition is expected because *again* can be ignored without making the sentence entirely senseless. For the definite determiner this is not possible because assertion and presupposition are dependent on each other. For scalar implicatures processing difficulties are expected depending on the theoretical approach, these difficulties are expected to appear in different conditions (a detailed discussion of different theoretical approaches for scalar implicatures can be found in the manuscript in the appendix). Following Bott and Noveck (2004) a slower and delayed processing of scalar implicatures is expected, whereas for presuppositions an early processing starting already on the trigger (see e.g., Bade & Schwarz, 2019a) is assumed.

As expected, the acceptability ratings were high in the verifying condition and quite low in the falsifying condition supporting the idea that participants perceived the inappropriateness of a context that explicitly falsified the presuppositional/implicated content of the test sentence. Reading times were analyzed as reading times per letter. The results revealed significant differences between the four sentence types for all analyzed positions. For the determiner sentences, longer reading times were observed in the falsifying condition compared to the verifying condition before the end of the sentence which suggests an immediate processing of presuppositions. A failure of this verification process results in additional processing costs. For the trigger *again* this difference between the two conditions could not be observed, which highlights that different triggers are processed differently and thus require separate analyses. For the implicatures longer reading times were observed in the verifying condition compared to the falsifying condition, which suggests that processing the implicature is more difficult than ignoring it in case it is not supported by the context. Those processing difficulties occurred quite early while reading the sentence and thus suggest that implicatures are immediately available (see also, Foppolo & Marelli, 2017). Although the processing of scalar implicatures also starts early, the processing differs from the processing of presuppositions because the effect of context condition persists until the end of the sentence. While the processing of scalar implicatures is a long-lasting process, presupposition processing seems to be completed before the end of the sentence.

A possible explanation for the absence of an effect between the verifying and the falsifying condition for the sentences with the trigger *again* could be that the verifying context was not sufficiently specific. More precisely, the presented context did not mention a specific event. It only said that, for example, Lukas has ordered pizza sometime before the utterance, but not exactly when this event took place, like in *Last Saturday, Lukas ordered pizza*…. This could complicate the situation, because the addressee struggles identifying the exact event which could lead to a failure of the verification process at the evaluation word. The current study provided evidence for an immediate verification process, for presuppositions triggered by definite determiners starting as soon as the content of the presupposition is known. In case of failure, this process evokes longer reading times than in a successful verification. According to the current data, this cannot be claimed for presuppositions triggered by *again*. This strengthens the claim that different presupposition triggers need to be analyzed separately. Although the processing of presuppositions and of scalar implicatures start quite early, the two processes differ because for scalar implicatures the effect of context condition persists till the end of the sentence. In summary, the evaluation process of scalar implicatures seems to be a long-lasting process which is not the case for presuppositions.

7. General Discussion

The question of how listeners process the meaning of an utterance in a certain context is important for the study of natural language. Speakers often communicate more than they actually say. Those additional meaning components can be distinguished from the pure assertion, but how do they arise? When does the listener process this additional meaning component? The present dissertation investigated the role of presuppositions with a focus on presuppositions triggered by definite determiners and contrasts this classical presupposition trigger with indefinite determiners. The mouse-tracking methodology provides evidence for a two-step processing of the anti-uniqueness inference of the indefinite determiner. Furthermore, the combination of classical self-paced reading experiments with the PRP paradigm from cognitive psychology made it possible to investigate the underlying processes and to gather evidence for a capacity-limited processing of presupposition triggers.

7.1 Summary of results

In Chapter 5, I formulated three research questions that were investigated in the present dissertation. Before discussing the results in detail, I will provide an overview of which study investigated those research questions and the preliminary answers that can be given according to the respective data. Table 2 sums up the results at a glance. The following sections focus on the three questions separately and discuss the results in more detail.

Table 2: Overview which study provides evidence to answer the three research questions.

	Q1: How does the anti-uniqueness inference of the indefinite determiner arise?	**Q2:** What cognitive resources are involved during presupposition processing?	**Q3**: When do presuppositions unfold their impact?
Study 1	Maximize Presupposition	capacity-limited	immediately
Study 2	Maximize Presupposition	capacity-limited	
Study 3		capacity-limited	immediately
Study 4	Maximize Presupposition		immediately
Study 5			immediately

7.1.1 How does the anti-uniqueness inference of the indefinite determiner arise?

The status of the definite determiner as a presupposition trigger, triggering existence and uniqueness, is undoubtedly clear. This is not the case for the indefinite determiner, which is associated with an anti-uniqueness inference.

Study 1 supports the anti-presupposition theory, because for all dependent variables larger values were observed for the indefinite compared to the definite determiner. Especially the X_{neg} results provide evidence for an initial activation of the uniqueness-presupposition of the definite determiner when processing the indefinite determiner. This result was only predicted by the anti-presupposition theory based on the Maximize Presupposition principle (Heim, 1991). Following this idea, anti-uniqueness is derived in a two-step process. First the stronger alternative of the definite determiner is considered, and secondly, the uniqueness-presupposition of the definite determiner is negated. Using mouse-tracking is a great opportunity to visualize this initial activation of the uniqueness-presupposition of the definite determiner. The mouse trajectory initially deviates towards the later not-selected response option, reflecting its brief activation. This deviation is reflected in X_{neg} scores and the data clearly revealed larger X_{neg} scores for the indefinite compared to the definite determiner. The presented data clearly supports the idea that the anti-uniqueness associated with the indefinite determiner is neither a presupposition nor an implicature, but results from the Maximize Presupposition principle.

Further evidence is provided by RT data of Study 2. RTs for the sentence evaluation task were longer for sentences including an indefinite determiner compared to sentences with a definite determiner. This again suggests that processing indefinite determiners is more difficult and thus may require an extra processing step compared to definite determiners. The application of the PRP paradigm helped to minimize potential candidates for this additional processing step. It revealed an additive combination with SOA, thus processing the indefinite determiner requires central capacities. As assumed by the Maximize Presupposition principle negation is involved during the processing of the indefinite determiner and negation has been suggested to be a capacity-limited process in previous research (e.g. Deutsch et al., 2009; Foerster et al., 2019; Strack & Deutsch, 2004; Wason, 1959).

Study 4 revealed that participants are aware of the additional meaning components of uniqueness and anti-uniqueness of the definite and indefinite determiner. Participants use those components to disambiguate sentences as early as possible as long as the determiners are used felicitously. The results regarding differences between the two determiners were less clear. The time-based measures revealed longer times for the indefinite compared to the definite determiner, supporting the anti-presupposition theory. However, no significant difference between the two determiners was observed for AUC, thus supporting the idea that both determiners trigger their own presuppositions proper. In this case, no processing difference would be expected from a theoretical point of view. Although not significant, the descriptive pattern still revealed larger AUC values for the indefinite compared to the definite determiner which is in line with the anti-presupposition theory.

One factor that may have caused those less clear results concerning the expected difference between the two determiners in Study 4 might be that definite and indefinite determiners were presented in a contrastive way. Previous work by Bade and Schwarz (2019a, 2019b) suggest that target choices for indefinite determiners are boosted when participants are also faced with alternatives (the definite determiner). For example, the production task administered before the mouse-tracking task in Study 4 helped to stabilize the anti-uniqueness inferences of the indefinite determiner, but this procedure makes it harder to distinguish between definite and indefinite determiners. It is a challenge to construct experiments in such a way that participants easily get the anti-uniqueness inference of the indefinite determiner without putting too much focus on it so that participants still interpret the determiners in a natural way. The presented experiments can function as a starting point, but further research is necessary.

7.1.2 What cognitive resources are involved during presupposition processing?

To achieve a deeper understanding of the processes involved in processing presuppositions and to further distinguish the (anti-)uniqueness inferences associated with the two determiners, I investigated capacity limitations during the processing of presuppositions triggered by definite and indefinite determiners.

To do so, in Study 2 and 3, I combined a reading task and the PRP approach and applied the locus of slack-logic. This procedure pursued the goal to evaluate whether presupposition processing is automatic and thus can run in parallel to other tasks or whether it is capacity-limited and has its locus within the central stage of processing. Study 2 compared the definite and the indefinite determiner. As suggested by the Maximize Presupposition principle, the longer RTs for the indefinite determiner are the result from negating the uniqueness-presupposition of the definite determiner. Previous research argued that negation is a capacity-limited process (e.g., Deutsch et al., 2009; Foerster et al., 2019; Strack & Deutsch, 2004;

Wason, 1959), and Study 2 revealed that one (or more) processes involved in processing the indefinite determiner requires limited cognitive capacities. Thus, the data is in line with the anti-presupposition theory that suggests that the anti-uniqueness inference associated with the indefinite determiner arises in a two-step process due to the Maximize Presupposition principle. The first step of the evaluation of the indefinite determiner is evaluating the uniqueness-presupposition of the definite determiner and secondly negating it. The knowledge of the process being capacity-limited reduces the number of possible candidates that can explain this procedure and negation seems a viable candidate, but Study 2 could not answer whether limited-cognitive capacities are required by other presupposition triggers (e.g., *again*) in comparison to neutral words. This was the focus of Study 3.

Study 3 focused on the immediate processing of presuppositions and investigated whether this process is automatic or capacity-limited. *Again* was included as another presupposition trigger. Similar to Study 2, the PRP approach was combined with a self-paced reading task. The central question of this study was, What kind of cognitive capacities are required for presupposition processing in comparison to the processing of neutral words? According to the data, automatic processing of presupposition triggers appears unlikely. When encountering a presupposition trigger a context search starts immediately. In case of the definite determiner, this is the search for an appropriate referent, for *again* a suitable previous event has to be found in the context. This search process likely involves repeated selection and deselection of working memory items which can be considered as possible referents or relevant previous events. According to Janczyk (2017), selecting working memory items is a capacity-limited process which can explain why the processing of presupposition triggers requires limited cognitive capacities.

The required capacities can provide further insight into the processing differences of different pragmatic phenomena. While it is assumed that presuppositions are processed immediately, Study 5 suggests that the processing of implicatures seem to be a long-lasting process. With the PRP paradigm and the locus of slack-logic the underlying processes that are involved in the processing of scalar implicatures can be investigated. As negation is also involved in the processing of implicatures (e.g. *some, but not all*), I expect that the processing is capacity-limited and not automatic. As the data in Study 5 suggested, processing scalar implicatures persists till the end of the sentence. Applying the PRP approach to different positions in the sentence might reveal where capacity-limited processes are involved and thus can narrow down when negation unfolds its impact. To investigate this question, several experiments are necessary because the PRP paradigm can only be applied to one position in a sentence in one experiment. An investigation of scalar implicatures with the presented paradigm could shed light on this interesting question in future research.

7.1.3 When do presuppositions unfold their impact?

Finally, I turn to the question dealing with the temporal dynamics of processing that asked: When do presuppositions unfold their impact? Previous results suggested an immediate processing of presuppositions starting already on the trigger (e.g., Bade & Schwarz, 2019b; Kirsten et al., 2014; Tiemann et al., 2011). This could also be concluded from the X_{neg} results in the mouse-tracking data of Study 1. Additional evidence was provided by the self-paced reading data of Experiment 1 of Study 3. The reading times of trigger words (definite determiner and *again*) were longer than the reading times of an unacceptable word at the relevant position. These results speak for an immediate processing of the presupposition trigger. The data of Study 4 also support this idea because it revealed early effects before the end of the sentence, which suggests that participants use the additional meaning component triggered by definite and the indefinite determiner as early as possible. Study 5 was dedicated to distinguish presuppositions that were explicitly verified by the context from presuppositions that were falsified by the context. The data also revealed effects as soon as the presuppositional content was known to the participant for presuppositions triggered by determiners. The data did not reveal this effect for the trigger *again*. Thus, Study 5 also supports the idea of immediate processing of presuppositions triggered by definite determiners. However, one has to point out that the additional meaning component of the definite and indefinite determiner in Study 4 were only used to disambiguate the sentence in case the determiners were used felicitously. In the unreliable group this effect was not observed. This suggests that participants stopped using the additional meaning component as they were faced with inappropriate or infelicitous uses. In case the context explicitly falsifies the presupposition, processing the definite determiner is prolonged, as can be seen in Study 5. This does not seem to be the case for the trigger *again*. One has to confess though, that the unclear results for the trigger *again* might be due to the used items. As discussed in the manuscript, it might be that the context used for the trigger *again* was not explicit enough. To clarify this suggestion further research is necessary.

7.2 Implications for theories of presuppositions

The definite determiner is a classic example for a presupposition trigger (e.g., Frege, 1892). The status of the additional meaning component of anti-uniqueness associated with the indefinite determiner is, however, controversially discussed in the literature. Three theories were discussed in this dissertation. First of all, as suggested in Kratzer (2005), the indefinite determiner presupposes anti-uniqueness and thus is a presupposition trigger, similar to the definite determiner (see also Heim, 1991, 2011 for a discussion). A second explanation goes back to Chierchia et al. (2012) and Grønn and Sæbø (2012) who treat the anti-uniqueness of the indefinite determiner as an implicature. Finally, the anti-uniqueness of the indefinite determiner can be explained as an anti-presupposition (Percus, 2006) resulting from pragmatic reasoning based on the Maximize Presupposition principle (Heim, 1991). Studies 1, 2, and 4 provided evidence for a more complex processing of the indefinite compared to the definite determiner and thus support the idea of anti-uniqueness of the indefinite determiner arising as an anti-presupposition. Especially Study 1 provided interesting insight into the processing of the indefinite determiner and suggests that this processing runs in two steps. The mouse-tracking data can make the initial activation of the uniqueness-presupposition of the definite determiner visible and thus provides experimental evidence for the psychological reality of the anti-presupposition theory.

Study 2 provided further evidence for this theory. According to the two-step model derived from the Maximize Presupposition principle the processing of the indefinite determiner requires as a first step the evaluation of the uniqueness-presupposition of the definite determiner, and its negation in a second step. Study 2 applied the locus of slack-logic and the data revealed additivity with SOA, which means that one (or more) processes involved in processing the indefinite determiner require limited cognitive capacities. Previous research (e.g., Deutsch et al., 2009; Foerster et al., 2019; Strack & Deutsch, 2004; Wason, 1959) assume that negation is a capacity-limited process. The presented results thus replicate previous findings of a more complex processing of the indefinite compared to the definite determiner (see, e.g., Bade & Schwarz, 2019a; Hertrich et al., 2015; Kirsten et al., 2014), and furthermore, it helps to delineate to nature of the additional process that causes longer RTs for the indefinite determiner. The two-step model with negation of the uniqueness-presupposition is a viable candidate to explain this difference. The presented experiments, thus clearly support the anti-presupposition theory and provide experimental evidence for the Maximize Presupposition principle.

Furthermore, the investigation of the implicated processes during presupposition processing revealed that those processes are capacity-limited in general (see Study 3) and thus cannot run in parallel with other ongoing processes. Those findings are in line with previous research that assumes that when a presupposition trigger is encountered an immediate context search for an appropriate referent is started. This search is likely to involve repeated selection and de-selection of working memory items which are potential candidates as referents. Janczyk (2017) showed that selecting working memory items is a capacity-limited process. Additionally, when the search process is successful, a referential assignment of the noun phrase to an antecedent is necessary. This referential assignment requires encoding of information in working memory, which has been shown to be capacity-limited (Jolicoeur & Dell'Acqua, 1998). The presented data thus supports previous findings on presupposition processing with data from another perspective. This account is also promising to shed light onto the still unclear processes that are required in case of accommodation.

However, the presented experiments also revealed that different presupposition triggers might require separate analyses, because the underlying processes differ in difficulty and it is possible that additional processes are involved only in processing some triggers. To generalize this future research with other presupposition triggers is necessary.

7.3 Limitations and future extensions

The probably clearest limitation of the presented work is the focus on only one language, namely German. Future research has to take other languages into account to clarify whether processing of presupposition triggers is generally a capacity-limited process. Furthermore, the dissertation mainly focused on one presupposition trigger, namely definite determiners. Study 3 and 5 already revealed that presupposition triggers likely differ in their processing and future research should consequently address other triggers as well. To further strengthen the idea of the Maximize Presupposition principle, "both" and "exactly 2" seem to be interesting candidates. Focusing only on one (or at most two) trigger(s) do(es) not allow to transfer the results to presupposition triggers in general, but it has the great advantage to increase the number of items which allows separate meaningful analysis. Study 5 tried to integrate other pragmatic phenomena like implicatures. The comparison of different pragmatic phenomena is an interesting and fruitful field as well. The integration of the PRP approach and the locus of slack-logic can be used to shed further light on the underlying processes and improve understanding the processing of scalar implicatures.

8. Conclusion

The main question of the present dissertation was how presuppositions are processed with a focus on the definite and the indefinite determiner. The data revealed that presuppositions are processed immediately (see also, Bade & Schwarz, 2019b; Burkhardt, 2006; Kirsten et al., 2014; Tiemann et al., 2011). Furthermore, the data provide experimental evidence for a more difficult processing in case of infelicitously used triggers and thus replicates previous results by Tiemann et al. (2011), (see also, Hertrich et al., 2015; Van Berkum et al.,1999; Van Berkum et al.,2003). In addition, the presented work tried to answer the question whether the definite and the indefinite determiner are both presupposition triggers. Data from study 1,2 and 4 are not in line with this assumption and thus, I do not regard the indefinite determiner as a presupposition trigger, but instead support the anti-presupposition theory based on the Maximize Presupposition principle. The experimental data provide evidence for a more complex processing of the indefinite compared to the definite determiner which results from a two-step processing of the indefinite determiner that includes the evaluation of the uniqueness-presupposition of the definite determiner in a first step and its negation in a second step. Furthermore, the application of the locus of slack-logic in PRP experiments provide further evidence for this theory because the data revealed that for the indefinite determiner a capacity-limited process is involved. Previous research on negation reflected that this is a capacity-limited process (see e.g., Deutsch et al., 2009; Foerster et al., 2019; Strack & Deutsch, 2004; Wason, 1959) and thus the anti-presupposition theory is strengthened. Comparing the processing of presupposition triggers and other neutral words revealed that the processing of presuppositions is a capacity-limited process which cannot run in parallel with other ongoing processes. Including another presupposition trigger made clear that different presupposition triggers require separate analyses to make adequate statements due to the underlying processes during the presupposition processing.

The dissertation brought together experimental pragmatics and cognitive psychologist's (chronometric) tools. This novel combination made it possible to investigate the underlying processes that are involved in presupposition processing and it revealed that presupposition processing requires limited cognitive capacities. It also demonstrates that interdisciplinary investigations are a fruitful enterprise.

9. References

Abbott, B. (2000). Presuppositions as nonassertions. *Journal of Pragmatics, 32*, 1419-1437.

Abrusán, M. (2011). Predicting the presuppositions of soft triggers. *Linguistics and Philosophy, 34,* 491-535.

Abusch, D. (2009). Presupposition Triggering from Alternatives. *Journal of Semantics, 27*, 37-80.

Altmann, G., & Steedman, M. (1988). Interaction with context during sentence processing. *Cognition, 30*, 191–238.

Anderson, J. E., & Holcomb, P. J. (2005). An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain and Language, 94,* 200–216.

Ayora, P., Peressotti, F., Alario, F.-X., Mulatti, C., Pluchino, P., Job, R., & Dell'Acqua, R. (2011). What phonological facilitation tells about semantic interference: A dual-task study. *Frontiers in Psychology, 2,* 57.

Bade, N., & Schwarz, F. (2019a). (in-)definites, (anti-)uniqueness, and uniqueness expectations. In R. P. Cooper (Ed.), *Proceedings of CogSci 2019.* Montreal, Canada.

Bade, N., & Schwarz, F. (2019b). An experimental investigation of antipresuppositions. In A. Creemers, & C. Richter (Eds.), *Proceedings of PLC 42* (pp. 29–40). Philadelphia, Pennsylvania: University of Pennsylvania Working Papers in Linguistics.

Beaver, D., & Condoravdi, C. (2003). A uniform analysis of "before" and "after". In R. B. Young, & Y. Zhou (Eds.), *Proceedings of Semantics and Linguistic Theory XIII* (pp. 37-54). Washington: University of Washington.

Beaver, D., & Geurts, B. (2012). Presupposition. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 3, pp. 2432-2459). Berlin: Mouton de Gruyter.

Beck, S. (2007). Quantifier dependent readings of anaphoric presuppositions. In U. Sauerland, & P. Stateva (Eds.), *Presupposition and Implicature in Compositional Semantics* (pp. 12-33). UK: Palgrave Macmillan.

Bill, C., Romoli, J., & Schwarz, F. (2018). Processing presuppositions and implicatures: Similarities and differences. *Frontiers in Communication, 3*,1-44.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437-457.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language, 98,* 159–168.

Chemla, E. (2008). An epistemic step for anti-presuppositions. *Journal of Semantics, 25,* 141-173.

Chemla, E. (2009). Presuppositions of quantified sentences: Experimental data. *Natural Language Semantics, 17*, 299-340.

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In K. von Heusinger, C. Maienborn, & P. Portner (Eds), *Semantics: An International Handbook of Natural Language Meaning* (Vol. 3, pp. 2297-2331). Berlin: Mouton de Gruyter.

Clifton, C. (2013). Situational context affects definiteness preferences: Accommodation of presuppositions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 487-501.

Cooper, R. (1983). *Quantification and syntactic theory.* Dordrecht: Reidel.

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture–word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review, 14,* 717–722

Delin, J. (1995). Presupposition and shared knowledge in it-clefts. *Language and Cognitive Processes, 10*, 97-120.

Deutsch, R., Kordts-Freudinger, R., Gawronski, B., & Strack, F. (2009). Fast and fragile. A new look at the automaticity of negation processing. *Experimental Psychology, 56,* 434-446.

Domaneschi, F., & DiPaola, S. (2018). The processing costs of presupposition accommodation. *Journal of Psycholinguistic Research*, *47,* 483-503.

Durst, M., & Janczyk, M. (2018). The motor locus of the no-go based backward crosstalk. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 1931-1946.

Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Implications for the central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1058-1079.

Foerster, A., Wirth, R., Berghoefer, F. L., Kunde, W., & Pfister, R. (2019). Capacity limitations of dishonesty. *Journal of Experimental Psychology: General, 148,* 943-961.

Foppolo, F., & Marelli, M. (2017). No delay for some inferences. *Journal of Semantics, 34,* 659–681.

Frazier, L. (1978). Theories of sentence processing. In J. Garfield (Ed.), *Modularity in knowledge representation and natural language processing* (pp. 493-522). San Diego, CA, USA: Bradford Books.

Freeman, J. B., & Ambady, N. (2010). Mousetracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*, 226–241.

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology, 2,* 59.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik, 100*, 25-50.

Geurts, B. (1994). *Presupposing*. Dissertation, Universität Osnabrück.

Geurts, B., & van der Sandt, R. (2004). Interpreting focus. *Theoretical Linguistics, 30*, 1-44.

Grice, H. P. (1975). Logic and Conversation. In P. Cole, J. L. Morgan *Syntax and Semantics, Vol. 3, Speech Acts* (Vol. 3, pp. 41–58). New York: Academic Press.

Grønn, A., & Sæbø, K. J. (2012). A, the, another: A game of same and different. *Journal of Logic, 21, 75-95.*

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13,* 512-521.

Heim, I. (1982). The Semantics of Definite and Indefinite Noun Phrases. Ph. D. thesis, University of Massachusetts.

Heim, I. (1991). Artikel und Definitheit. In A. v Stechow, & D. Wunderlich (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft* (Bd. 6 Semantik, pp. 487-535). Berlin: De Gruyter.

Heim, I. (2011). Definiteness and indefiniteness. In C. Maienborn, K. von Heusinger, & P. Portner (Eds), *Semantics: An international handbook of natural language meaning* (pp. 996-1025). Berlin: De Gruyter Mouton.

Hendricks, M. A., Conway, C. M., & Kellogg, R. T. (2013). Using dual-task methodology to dissociate automatic from nonautomatic processes involved in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 39,* 1491-1500.

Hertrich, I., Kirsten, M., Tiemann, S., Beck, S., Wühle, A., Ackermann, H., & Rolke, B. (2015). Context dependent impact of presuppositions on early magnetic brain responses during speech perception. *Brain and Language*, *149,* 1-12.

Horn, L. R. (1972). *On the semantic properties of logical operators in English.* Dissertation, University of California.

Janczyk, M. (2013). Level 2 perspective taking entails two processes: Evidence from PRP experiments. In *Journal of Experimental Psychology: Learning, Memory and Cognition, 39,* 1878-1887.

Janczyk, M. (2017). A common capacity limitation for response and item selection in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1690–1698.

Janczyk, M., Humphreys, G.W., & Sui, J. (2019). The central locus of self-prioritization. *The Quarterley Journal of Experimental Psychology, 72*, 1068-1083.

Jolicoeur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36,* 138–202.

Jouravlev, O., Stearns, L., Bergen, L., Eddy, M., Gibson, E., & Fedorenko, E. (2016). Processing temporal presuppositions: An event-related potential study. *Language, Cognition and Neuroscience, 31,* 1245-1256.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111,* 228-238.

Karttunen, L. (1973). Presuppositions of compound sentences. *Linguistic Inquiry, 4*, 167-193.

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods, 49,* 1652-1667

Kiparsky, P., & Kiparsky, C. (1970). Fact. In M. Bierwisch, & K. Heidolph (Eds.), *Progress in Linguistics* (pp. 143-173), The Hague: Mouton.

Kirsten, M., Tiemann, S., Seibold, V. C., Hertrich, I., Beck, S., & Rolke, B. (2014). When the polar bear encounters many polar bears: Event-related potential context effects evoked by uniqueness failure. *Language, Cognition and Neuroscience, 29*, 1147–1162.

Kratzer, A. (2005). Building resultatives. In C. Maienborn, & C. Wöllenstein-Leisten (Eds.), *Event arguments in syntax, semantics, and discourse* (pp. 177-212). Tübingen: Niemeyer.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23-49.

Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistic electrified II: 1994- 2005. In M. Traxler, & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659-724). New York: Elsevier.

Langendoen, D. T., & H. Savin (1971). The projection problem for presuppositions. *Studies in Linguistic Semantics*. 55-60

Levinson, S. C. (1983). *Pragmatics*. Cambridge, UK: University Press.

Levy, J., & Pashler, H. (2001). Is dual-task slowing instruction dependent? Journal of *Experimental Psychology: Human Perception and Performance, 27,* 862-869.

Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy tradeoffs (?). *Behavior Research Methods, 51*, 40–60.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review, 108,* 393-343.

Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, *17*, 245–297.

Miller, J., & Reynolds, A. (2003). The locus of redundant-targets and nontargets effects: Evidence from the psychological refractory period paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 1126–1142.

Murphy, G. (1984). Establishing and accessing referents in discourse. *Memory & Cognition, 12*, 489-497.

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology*, *44*, 193-251.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 358–377.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220–244.

Percus, O. (2006). Antipresuppositions. In A. Ueyama (Ed.), *Theoretical and empirical studies of reference and anaphora: Toward the establishment of generative grammar as an empirical science* (pp. 52-73). Tokyo, Japan: Japan Society for the Promotion of Science.

Piai, V., Roelofs, A., & Schriefers, H. (2014). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 147-165.

Rayner, K., & Clifton, C. Jr. (2002). Language Processing. In D. Medin (Ed.), *Stevens Handbook of Experimental Psychology* (3rd ed., Vol. 2, pp. 261-316). New York, NY: John Wiley and Sons.

Roettger, T. B., & Franke, M. (2019). Evidential strength of intonational cues and rational adaptation to (un-)reliable intonation. In Psychological Research R. P. *Cooper Cognitive Science*, 43 (pp. e12745). Hoboken, NJ: Cognitive Science Society.

Russell, B. (1905). On denoting. *Mind, 14*, 479-493.

Sauerland, U. (2008). Implicated presuppositions. In A. Steube (Ed.), Sentence and Context. Language, Context & Cognition., Mouton de Gruyter.

Schneider, C., Schonard, C., Franke, M., Jäger, G., & Janczyk, M. (2019). Pragmatic processing: An investigation of the (anti-)presupposition of determiners using mouse-tracking. *Cognition, 193*, 104024

Schneider, C., Bade, N., Franke, M., & Janczyk, M. (2020). Presuppositions of determiners are immediately used to disambiguate utterance meaning. *Psychological Research,* https://doi.org/10.1007/s00426-020-01302-7.

Schneider, C., Bade, N., & Janczyk, M. (2020). Is immediate processing of presupposition triggers automatic or capacity-limited? A combination of the PRP approach with a self-paced reading task. *Journal of Psycholinguistic Research, 49,* 247-273.

Schneider, C., & Janczyk, M. (accepted). Capacity limitations of processing presuppositions triggered by determiners. *Acta Psychologica.*

Schwarz, F., & Tiemann, S. (2012). Presupposition Processing - The Case of German wieder. In M. Aloni, V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz, & M. Westera (Eds.), *Logic, Language and Meaning: 18th Amsterdam Colloquium* (pp. 200-209). Berlin: Springer-Verlag

Schwarz, F., & Tiemann, S. (2017). Presupposition Projection in Online Processing. *Journal of Semantics*, *34*, 61–106.

Schwarz, F. (2007). Processing presupposed content. *Journal of Semantics, 24*, 373- 416

Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a Stroop task. *Journal of Mathematical Psychology, 18*, 105-139.

Simons, M. (2001). On the conversational basis of some presuppositions. In R. Hastings, B. Jackson, & Z. Zvolenszky (Eds.), *Proceedings of Semantics and Linguistics Theory 11* (pp. 431-448). New York, NY: New York University.

Song, J. H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences, 13,* 360-366.

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic, 2,* 447–457.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220-247.

Strawson, P. (1950). On referring. *Mind, 59*, 320-344.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology, 14,* 1-36.

Thomason, R. (1972). A semantic theory of sortal incorrectness. *Journal of Philosophical Logic 1*, 209-258.

Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., Knapp, J., & Beck, S. (2011). Psycholinguistic evidence for presuppositions: On-line and off-line data. In I. Reich, E. Horch, & D. Pauly (Eds.), *Proceedings of Sinn & Bedeutung 15* (pp. 581–595). Saarbrücken: Saarland University Press.

Tombu, M., & Jolicœur, P. (2004). Virtually no evidence for virtually perfect time sharing. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 795-810.

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 3-18.

Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language, 69*, 18–35.

Tonhauser, J., Beaver, D. I., & Degen, J. (2018). How Projective is Projective Content? Gradience in Projectivity and At-issueness. *Journal of Semantics*, *35*, 495–542.

Van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*, 235–248.

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147–182.

Van de Meerendonk, N., Kolk, H. H. J., Vissers, C. Th. W. M., & Chwilla, D. J. (2008). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience, 22*, 67-82.

Van der Sandt, R. (1992). Presupposition projection as anaphora. *Journal of Semantics, 9*, 333-377.

Von Fintel, K. (2004). Would you believe it? The king of France is back! Presuppositions and truth value intuitions. In M. Reimer, & A. Bezuidenhout (Eds.), *Descriptions and beyond* (pp. 315-341). Oxford, UK: University Press.

Von Stechow, A. (1996). The different readings of Wieder "again": A structural account. *Journal of Semantics, 13*, 87-138.

Wason, P. C. (1959). The Processing of Positive and Negative Information. *Quarterly Journal of Experimental Psychology*, *11*, 92–107.

Welford, A. T. (1952). The ‘psychological refractory period’ and the timing of high-speed performance--A review and a theory. *British Journal of Psychology. General Section*, *43*, 2–19.

9.1 Reused material

Figure 4 and Figure 5 are published in the following article and were reused for the present dissertation:

Schneider, C., Schonard, C., Franke, M., Jäger, G., & Janczyk, M. (2019) Pragmatic processing: An investigation of the (anti-)presupposition of determiners using mouse-tracking. *Cognition, 193,* 104024.

Figure 6 is included in the following manuscript which is accepted for publication and was reused for the present dissertation:

Schneider, C., & Janczyk, M. (accepted). Capacity limitations of processing presuppositions triggered by determiners. *Acta Psychologica.*

Figure 7 is published in the following article and is reused for the present dissertation:

Schneider, C. Bade, N., Franke, M., & Janczyk, M. (2020). Presuppositions of determiners are immediately used to disambiguate utterance meaning. *Psychological Research.* https://doi.org/10.1007/s00426-020-01302-7.

10. Appendix

10.1 Study 1

Official citation:

Schneider, C., Schonard, C., Franke, M., Jäger, G., & Janczyk, M. (2019). Pragmatic processing: An investigation of the (anti-)presuppositions of determiners using mouse-tracking. *Cognition, 193,* 104024.

Cognition 193 (2019) 104024

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Pragmatic processing: An investigation of the (anti-)presuppositions of determiners using mouse-tracking

Cosima Schneider[a,*], Carolin Schonard[a], Michael Franke[b], Gerhard Jäger[c], Markus Janczyk[d]

[a] Department of Psychology, University of Tübingen, Tübingen, Germany
[b] Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany
[c] Department of Linguistics, University of Tübingen, Tübingen, Germany
[d] Department of Psychology, University of Bremen, Bremen, Germany

ARTICLE INFO

Keywords:
Presupposition
Anti-presupposition
Maximize Presupposition
Experimental pragmatics
Indefinite determiner
Mouse-tracking

ABSTRACT

A presupposition is a condition that has to be met in order for a linguistic expression to be appropriate. The definite determiner (as in *the banana)* triggers the uniqueness-presupposition that there is a uniquely identifiable banana in the relevant discourse context. The indefinite determiner (as in *a banana*) is similarly associated with anti-uniqueness (that there are several bananas). Application of the Maximize Presupposition principle to the indefinite determiner suggests that this latter effect results indirectly as an anti-presupposition from considering the uniqueness-presupposition of the definite determiner, which is then negated. This results in increased processing difficulty. We utilized mouse-tracking to compare processing of definite and indefinite determiners when used felicitously and infelicitously in a particular context. First, processing of the indefinite determiner was associated with more processing difficulties compared with the definite determiner. Second, we also observed evidence for an initial temporary activation and evaluation of the uniqueness-presupposition, just as derived from anti-presupposition theory and the Maximize Presupposition principle.

1. Introduction

How listeners grasp the overall conveyed meaning of an utterance in a certain context is an important question in the study of natural language meaning. Previous work in philosophy of language and linguistic semantics has put forward that several levels of meaning can be distinguished. On the one hand, there are the literal truth conditions, on the other hand, there are presuppositions and conversational implicatures. The present paper focuses on presuppositions—a vital topic in the semantic and pragmatic literature throughout the last decades (see Beaver, 2012). Specifically, we investigate the interpretation of definite and indefinite determiners by means of two mouse-tracking experiments. The results provide novel empirical evidence for a more difficult processing of indefinite determiners and their anti-presupposition, which is derived from a negation of the uniqueness-presupposition of the definite determiner, as predicted by theoretical accounts that rely on the *Maximize Presupposition* principle (Heim, 1991). The article is structured as follows: We begin by introducing presuppositions in general and those of determiners in particular (see Hawkins, 1978, for an overview or the discussion in Heim, 1991, 2011), including a brief presentation of general predictions derived from different theoretical accounts. The next section reviews previous experimental investigations of presuppositional content relevant to the current research. This is followed by an introduction of the methodology of mouse-tracking and the experimental task we use, plus the predictions derived from the adopted theoretical framework of anti-presupposition theory.

1.1. The presuppositions of determiners

A presupposition is a condition that has to be met in order for a linguistic expression to be appropriate (Heim, 1991; Heim & Kratzer, 1998; Stalnaker, 1973). As an illustration, consider the following examples:

(1a)	Context: There is one pen and one pencil on the desk. Please hand me the pen.
(1b)	Context: There are three pens and one pencil on the desk. # Please hand me the pen.

The examples in (1a) and (1b) show that certain linguistic expressions trigger appropriateness conditions, that is, the presupposition; the

* Corresponding author at: Department of Psychology, University of Tübingen, Schleichstraße 4, 72076 Tübingen, Germany.
E-mail address: cosima.schneider@uni-tuebingen.de (C. Schneider).

https://doi.org/10.1016/j.cognition.2019.104024
Received 15 November 2018; Received in revised form 26 June 2019; Accepted 7 July 2019

expression imposing the appropriateness condition is called a *presupposition trigger*.

1.1.1. The definite determiner

In Example (1), the definite determiner "the" triggers two presuppositions: that (i) there exists an entity of the type denoted by the noun phrases ("pen") and (ii) that this entity is unique (exactly one pen). Thus, the triggered presuppositions are referred to as *existence* and *uniqueness*. Because of the uniqueness-presupposition in Example (1), using the definite determiner is appropriate, that is, felicitous, against the context provided in (1a). In contrast, with the context provided in (1b), the uniqueness-presupposition is not met, and consequently the sentence is inappropriate, that is, infelicitous (rather than true or false; Heim & Kratzer, 1998).

Importantly, presuppositions are context dependent and their interpretation varies according to the domain of discourse in which they are interpreted. Generally, presuppositions have to be evaluated against the background of knowledge or beliefs shared by the speaker and the audience. Therefore, the domain of discourse in which a presupposition is evaluated can be just a small subset of the overall existing entities in the world, given by what speaker and listeners currently attend to as the relevant domain of discourse.

1.1.2. The indefinite determiner

The definite determiner has a competitor, namely the indefinite determiner, which seems to have its own felicity conditions. The examples in (2) and (3) show that the indefinite determiner is associated with anti-uniqueness (see also Heim, 2011; Percus, 2006), that is, the reverse of what is introduced by the definite determiner.[1] Accordingly, using the indefinite determiner is infelicitous if it is known that there is exactly one entity of the described kind in the relevant context of discourse, as in the Examples (2) and (3). For instance, in Example (2) "a sun" presupposes that there is more than one sun which conflicts with our common knowledge.

(2)	# A sun is shining.	
(3)	# A weight of our tent is under 4 lbs.	(Heim, 1991)

However, Heim (1991) observed that—under certain circumstances—it is possible to utter the sentence in Example (4) felicitously without being certain that there are at least two 20 feet long catfish:

(4)	Robert caught a 20ft. long catfish.	(Heim, 1991)

The sentence in this example is indeed appropriate if the speaker is not sure that there is exactly one catfish of that size. To account for the felicity of Example (4) and also the typical anti-uniqueness effect, Heim (1991) proposed adding another principle to Grice's (1975) conversational maxims: *Presuppose as much as possible!* (see also Chemla, 2009; and, for further discussions, Alonso-Ovalle, Menéndez-Benito, & Schwarz, 2011; Percus, 2006; Sauerland, 2008; Schlenker, 2012). More precisely, this Maximize Presupposition principle requires speakers to always use the felicitous sentence with the strongest presupposition among a set of alternatives, if the speaker knows that these presuppositions are fulfilled. In this way, Maximize Presupposition accounts for the fact that indefinite determiners of the form "a X" (where X is a noun phrase) are infelicitous when it is common ground that there is exactly "one X", such as in Examples (2) and (3). The listener in Example (3), for instance, assumes that the speaker wants to convey that the presupposition of the stronger alternative (i.e., that there is a unique weight of the tent) is not met, since the speaker did not obey Maximize Presupposition. Because this contrasts with common knowledge (everything has exactly one weight), it therefore leads to the oddness of (3). Inferences that result from pragmatic reasoning based on Maximize Presupposition have special properties and thus must be distinguished from literal semantic meaning, ordinary presuppositions, and conversational implicatures: They are weak inferences which can be cancelled, similar to implicatures, but different from semantic meaning and presuppositions. Further, similar to presuppositions, they are projective content (Sauerland, 2008), meaning that they survive when embedded in, for example, negations or questions. As a consequence, they have been given special names; we will adopt the term "anti-presupposition" (Percus, 2006) here and speak of the "anti-presupposition theory" of indefinite determiners in the following.

In sum, Maximize Presupposition accounts for anti-uniqueness via preventing use of a sentence (sometimes referred to as 'blocking a sentence') by a cooperative speaker, when there is a competing sentence that presupposes more and otherwise communicates the same. For example, the definite determiner "the" is presuppositionally stronger than the indefinite determiner "a", because "the" triggers existence and uniqueness, whereas "a" does not presuppose anything. In situations where there is exactly one object of a specific kind, it is then felicitous to use the definite determiner and Maximize Presupposition blocks the use of the indefinite determiner. In contrast, when the indefinite determiner is used, it can be reasoned that the speaker was not in a position to use the definite determiner instead.

1.1.3. Processing predictions from anti-presupposition theory

The previous section introduced the indefinite determiner as a competitor of the definite determiner. However, it is an open question, what the relevant cognitive processes are by which the meanings of determiners are computed and evaluated. Current theories suggest two different possibilities for how the anti-uniqueness of the indefinite determiner arises: On the one hand, if we do not adopt the anti-presupposition theory of indefinite determiners, but instead assume that both definite and indefinite determiners presuppose (in a standard sense) uniqueness and anti-uniqueness, respectively (see Kratzer, 2005, and the related discussions in Heim, 1991, 2011), only very weak predictions are possible: Their processing might in fact be different, but it is difficult, if not impossible, to predict which determiner is more difficult to process. Thus, predicting no processing difference between definite and indefinite determiners would be most straightforward.

On the other hand, when adopting the anti-presupposition theory of indefinite determiners based on the Maximize Presupposition principle, more precise and testable predictions can be formulated. Specifically, while a sentence with a definite determiner presupposes uniqueness, the indefinite determiner leads to anti-uniqueness only indirectly, following a two-step model: (1) the evaluation of the uniqueness-presupposition of the definite determiner and (2) the subsequent negation of exactly this presupposition. In other words, anti-presupposition theory predicts that encountering an indefinite determiner is associated with additional processing costs compared to the evaluation of the definite determiner through the additional step of negating the uniqueness-presupposition. Because of the first step, we expect the uniqueness-presupposition to be activated initially (at least temporarily), but it is important to note that the two steps proceed serially and the uniqueness- and the anti-uniqueness consideration are not active at the same time. As will become obvious in a later section, this approach makes surprising predictions, and here we aim to provide experimental evidence for them. We will now continue with presenting previous work speaking to this issue.

[1] Heim (1983, 2011) also points out that the competition between the definite and the indefinite determiner can also invoke familiarity versus novelty. More precisely, the indefinite determiner can also be used to introduce an entirely novel object to the current context. A more thorough discussion of this is beyond the scope of the present paper, though it has to be kept in mind when designing experiments in a way that put anti-uniqueness into focus, while preventing issues of novelty as much as possible.

1.2. Previous experimental investigations of presuppositional content

In this section, we first summarize previous studies on the processing of determiners. The section closes with a short summary of work in the field of scalar implicatures, largely because these studies employed a similar methodology as we will use in our experiments.

By now, many studies have investigated the processing of presuppositions (for a recent review, see Schwarz, 2016a). Regarding definite and indefinite determiners, one of the first studies was run by Murphy (1984), in which participants read stories sentence by sentence. Experiment 1 focused on the novelty aspect of indefinite determiners and compared noun phrases in contexts that only differed in the use of the definite or indefinite determiner (e.g., "Later, George was passed by a/the truck, too"). Reading times were significantly longer for the indefinite compared to the definite determiner, supporting the idea that finding an antecedent for a definite reference is easier than establishing a new referent in the discourse model. Experiment 2 investigated the difference between definite and indefinite references by using pronouns (e.g., "Sandy rode a bicycle and I rode it, too." vs. "Sandy rode a bicycle and I rode one, too.") and contrasted the indefinite pronoun "one" with the appropriate definite pronoun. The effect of Experiment 1 was replicated. (Experiments 3 and 4 demonstrated further that sentences comprising singular items are read faster than those with plural items.)

Altmann and Steedman (1988) ran a self-paced reading experiment and demonstrated processing costs associated with violating the presuppositions of determiners, focusing on the uniqueness-presupposition of the definite determiner. Therefore, they presented test sentences in two different contexts: Context 1 introduced two candidates for the referent (e.g., a safe with a new lock and a safe with an old lock), while Context 2 introduced exactly one candidate as the referent:

(5a) Context 1: A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was a safe with a new lock and a safe with an old lock.
(5b) Context 2: A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was a safe with a new lock and a strongbox with an old lock.

In the test sentence in (6), the prepositional phrase *with the new lock* modified *the safe* and therefore the uniqueness-presupposition is met, even when the sentence is presented in a context like (5a). This is, however, not the case for the test sentence in (7).

(6) The burglar/blew open/the safe/with the new lock/and made of/with the loot.
(7) The burglar/blew open/the safe/with the dynamite/and made of/with the loot.

The results of this experiment revealed larger processing costs during the disambiguation region (i.e., the underlined prepositional phrase in (6) and (7)), when the uniqueness-presupposition of the definite determiner is not met. Thus, people experience these processing difficulties already before the end of the sentence. A similar conclusion results from the studies by van Berkum, Brown, and Hagoort (1999) with written language and van Berkum, Brown, Hagoort, and Zwitserlood (2003) with spoken language. In both studies, event related potentials (ERP) of the EEG were analyzed and the definite determiner evoked early ERP effects when the uniqueness-presupposition was not met (for further evidence, see Burkhardt, 2006; Haviland & Clark, 1974; Schumacher, 2009).

Tiemann et al. (2011) reported evidence for an immediate processing of presuppositions and longer reading times in falsifying versus supporting contexts from three acceptability ratings and self-paced reading studies on different presupposition triggers (German "wieder" (English "again"), "auch" (English "also"), "aufhören" (English "stop"), "wissen" (English "know'"), and definites in the shape of possessive noun phrases). First, sentences with an unfulfilled presupposition, that is, an infelicitous use of the presupposition trigger, were rated as less acceptable than sentences without a presupposition trigger, but more acceptable than ungrammatical sentences. Second, reading times for the word following the trigger in sentences with presuppositions were longer than in the corresponding grammatical condition without a presupposition, but shorter than in the ungrammatical sentences. Thus, an evaluation of a presupposition seems to be initiated upon encountering the trigger. Third, when the authors compared sentences with a presupposition when the given context explicitly verified versus falsified the presupposition, reading times were significantly longer in the falsifying than in the verifying condition in the region where the content of the presupposition was known (i.e., in the disambiguation region). Because reading time analyses did not distinguish the different presupposition triggers, it remains unclear though whether the results are valid for all triggers or were driven by a subset of them.

The studies mentioned so far focused on, or at least included, the definite determiner. Other studies directly compared the definite with the indefinite determiner and provided evidence that readers are aware of the semantics and pragmatics of (in)definite determiners and associate the definite determiner with uniqueness and the indefinite determiner with anti-uniqueness. For example, two experiments of Clifton (2013) reported more processing difficulties for indefinite than for definite determiner phrases. Participants were presented with context sentences that established whether there are one or multiple of the relevant items (e.g., "In the kitchen …" vs. "In the appliance store…") and the following test sentence contained either a definite or an indefinite determiner ("The stove…" vs. "A stove…"). Reading times in contexts that stereotypically provided a single possible referent for the definite determiner phrase or multiple possible referents for an indefinite determiner phrase were shorter than when context and type of determiner did not match (e.g., definite determiner in a context providing multiple items). It is noteworthy that these results were only obtained when participants were required to perform an additional simple arithmetic task in between reading the sentence and answering a question about it (Exp. 2 and 4). By this additional task, Clifton forced the participants to comprehend the sentence to a deeper level, because he prevented rehearsing the sentence until the question was presented.

Kirsten et al. (2014) also investigated the uniqueness-presupposition of the definite determiner and the anti-uniqueness associated with the indefinite determiner. The authors constructed sets consisting of two types of context sentences (see Examples (8a) and (8b)) and two types of test sentences that were similar except for the used determiner ("the/a"; see Example (9)). Both test sentences were combined with both possible contexts: Using the definite determiner in a test sentence matched with the context in Example (8a), but mismatched with the context in Example (8b); in contrast, using the indefinite determiner matched with the context in Example (8b), but mismatched with the context in Example (8a).

(8a) Antje visited the Duesseldorf zoo yesterday and saw a polar bear in the bear enclosure.
(8b) Antje visited the Duesseldorf zoo yesterday and saw some polar bears in the bear enclosure.
(9) Antje noticed that the/a polar bear was very aggressive.

In their self-paced reading experiment, Kirsten et al. (2014) recorded EEG and observed processing differences between both determiners in ERP amplitudes but not in their latencies. In addition, the data indicate similar processing costs for both determiners in mismatching (i.e., in infelicitous) sentences, where the uniqueness/anti-uniqueness condition of the determiner is not met. Nonetheless, the authors suggested qualitatively different underlying processes for both determiners, for example, an increased impact on working memory in case of the definite determiner. Similar results were obtained in a magnetoencephalography (MEG) study with auditorily presented test sentences (Hertrich et al., 2015).

In sum, the reviewed studies suggest that a presupposition trigger is processed immediately or briefly after encountering it (e.g., Kirsten et al., 2014; Schwarz, 2007; Tiemann et al., 2011). Furthermore, processing seems to be more difficult when the presupposition is not met by the context, that is, in infelicitous compared with felicitous sentences (e.g., Burkhardt, 2006; Kirsten et al., 2014; Tiemann et al., 2011), although this may depend on additional cognitive load (Clifton, 2013). Finally, processing of the definite and indefinite determiner appears to differ (e.g., Kirsten et al., 2014; van Berkum et al., 1999, 2003), but experimental studies investigating the time-course and the reasons for these differences are rare. Importantly, to the best of our knowledge, no study so far directly demonstrates that the indefinite determiner indeed implicates the initial activation of the uniqueness-presupposition of the definite determiner, as is suggested by anti-presupposition theory. From a cognitive psychologist's point of view, one would expect that the assumed activation should leave measurable traces on human behavior; this is what the present study aims to demonstrate.

We will close this section by briefly turning to the field of scalar implicatures where two-step models are also discussed. These studies provide the methodological background for our experiments, although it is important to note that the two processing steps for scalar implicatures differ from those we suggest for the indefinite determiner: Implicatures involve the pragmatic enrichment of truth conditional semantic content, while anti-uniqueness of the indefinite determiner involves negation of presuppositional semantic content. Nevertheless, we expect those underlying processes to leave similar traces in the mouse-tracking data. For instance, Tomlinson, Bailey, and Bott (2013) used mouse-tracking to distinguish a one-step model, where the implicature is evaluated directly in only one step, and the two-step model of processing implicatures (see also Bott & Noveck, 2004). Participants were asked to judge sentences like "Some elephants are mammals "as "true" or "false". Importantly, such sentences can be evaluated in a logical way ("some and possibly all") suggesting a "true" response, or in a pragmatic way ("some, but not all") suggesting a "false" response. Participants were to move the mouse cursor to response boxes in the two top corners labeled as "true" or "false", and the two-step model predicts initial deviations toward the "true" response before then changing direction toward the "false" response for the pragmatic, but for the logical, interpretation. For example, in Experiment 1, Tomlinson et al. used a training phase with error feedback (see also Bott & Noveck, 2004) to bias participants into interpreting the sentences in either the logical or the pragmatic way. The data revealed the predicted pattern, and this result supports the suggested two-step processing of scalar implicatures, that is, people first access a truth conditional semantic content meaning and then enrich this meaning pragmatically with the scalar implicature (Tomlinson et al., 2013). Similar results were obtained by Huang and Snedeker (2011) in an eye-tracking study, further corroborating this conclusion.

In sum, these experiments showed that the methodology of mouse-tracking is suited to demonstrate evidence for a two-step model involved in processing scalar implicatures. In the present study, we adopted this methodology to investigate a two-step model of processing the indefinite determiner. Importantly though, the two steps suggested for implicatures and anti-presuppositions are different ones: The two steps for processing scalar implicatures are (1) evaluation of the truth conditional semantic content meaning, followed by (2) the pragmatic enrichment, whereas for the indefinite determiner they are (1) consideration of the uniqueness-presupposition of the definite determiner and (2) the negation of this presuppositional content. Hence, although both are pragmatic phenomena, the assumed processing steps clearly differ.

1.3. The present study

We compared the unfolding of processing the interpretation of definite and indefinite determiners in two mouse-tracking experiments. In the following, we first introduce this particular method, then we explain the setup and idea of our two experiments, and finally relate our hypotheses to the variables obtained from mouse-tracking.

1.3.1. Mouse-tracking as a method

Hand movements have been shown to be very sensitive to reveal the temporal dynamics of cognitive states (Song & Nakayama, 2009; see also Freeman, Dale, & Farmer, 2011). Recording movements of the computer mouse on its way to a particular goal-location representing a certain decision is an easy way to gather the required continuous data, and has become increasingly popular in cognitive psychology and experimental linguistics in recent years.

One of the first applications of this method to language processing used mouse movements to probe whether phonologically similar words are activated during sentence processing (Spivey, Grosjean, & Knoblich, 2005). Participants were presented with one picture in the upper left and another one in the upper right corner of the screen. In one condition, the depicted words were similar in their initial phonemes (e.g., "candle" vs. "candy"), while in another condition they were not (e.g., "candle" vs. "summer"). Participants started a mouse movement in the lower center of the screen and listened to an auditorily presented stimulus sentence, for example, "Click the candle!". When the initial phonemes of the words were similar, the trajectories of the movements towards the correct upper corner showed an attraction of the competitor. Thus, while processing the target word, competing phonological representations appear active and influence the exact way the hand moves. Similar approaches have since been applied to, for example, social cognitive questions (Freeman & Ambady, 2011), conflict tasks (Scherbaum, Dshemuchadse, Fischer, & Goschke, 2010), the effects of irrelevant stimulus variation on action execution (Janczyk, Pfister, & Kunde, 2013), or the influence of action consequences on action execution (Pfister, Janczyk, Wirth, Dignath, & Kunde, 2014; Wirth, Pfister, Janczyk, & Kunde, 2015). Mouse tracking has also been used to study conversational implicatures (Sauerland, Tamura, Koizumi, & Tomlinson, 2015; Tomlinson & Bott, 2013), predictive disambiguation based on early intonational cues (Roettger & Franke, 2018; Roettger & Stoeber, 2017), and sentence negation (Dale & Duran, 2011). Here, we apply this method to presupposition processing with the particular aim of providing evidence for the two-step processing of indefinite determiners.

While the mere trajectories already provide information about ongoing cognitive processes, several parameters are usually extracted from the trajectories for further statistical analyses. We here focus on the following parameters: (1) Area under the curve is the area between the empirical trajectory and the ideal path from the starting point to the end point of the trajectory, and is thought to provide a general measure of processing difficulty (see Freeman & Ambady, 2010): the more difficult the task the larger becomes the area under the curve. (2) Movement time is the time from stimulus onset until the cursor hits a response box. (3) X_{neg} is the amount of horizontal deviation toward the competitor (i.e., toward the non-target response) and is measured as the maximum x-coordinate reached by the mouse as it veers away from the target response. This measure is particularly useful for detecting two-step processes, in which the second step involves a negation of the first step and is therefore expected to trigger a reversal in mouse direction (cf. Tomlinson et al., 2013, for an example with implicatures).

1.3.2. General approach and procedure

The mouse-tracking task we used in our experiments is illustrated in Fig. 1. Participants were asked to judge the appropriateness of test sentences in the context of a short visualized story (see Fig. 2 for examples). Each trial of the experiment began with a context (see Fig. 1.1)

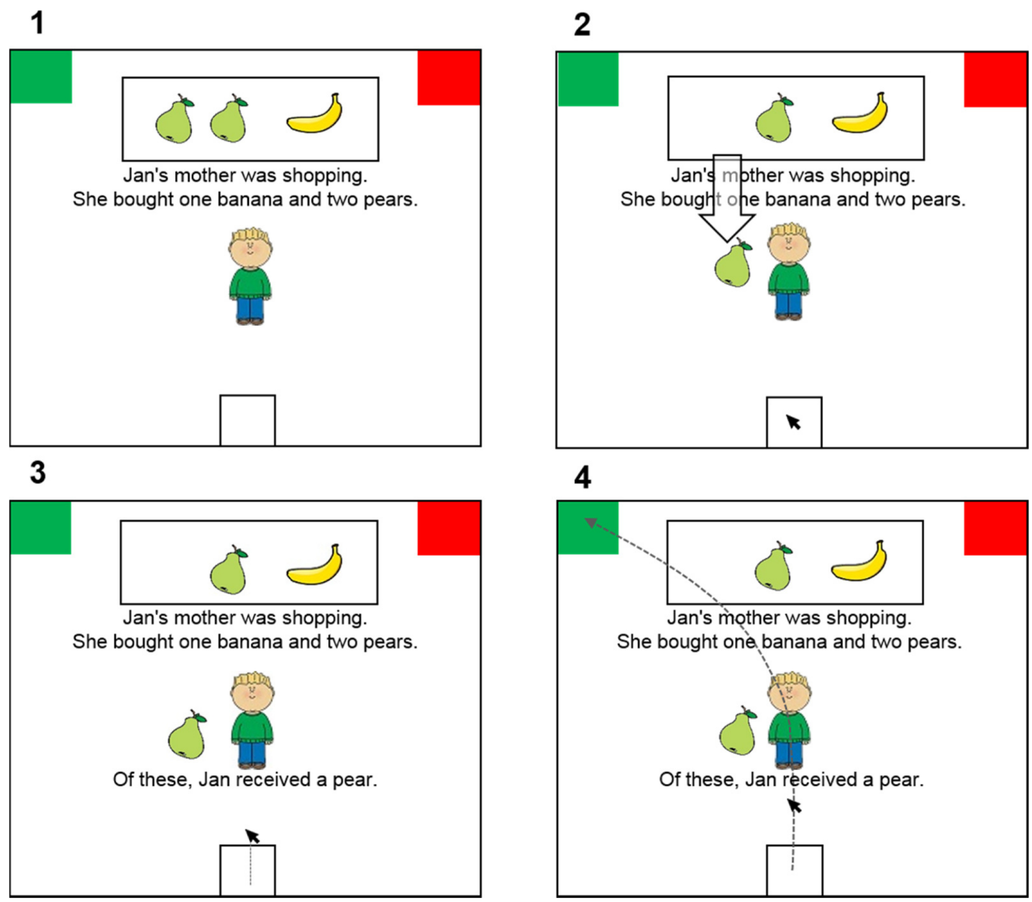


Fig. 1. Illustration of the task we used in the two experiments (see text for more information).

which contained a box (the "shopping basket") with three pieces of fruit in the upper part of the screen (the context relevant fruit) and was described with two sentences, for example, "Jan's mother was shopping. She bought one banana and two pears." (German original: "Jan's Mutter war einkaufen. Sie hat eine Banane und zwei Birnen gekauft."). Below was a picture of Jan. Participants were then to move the mouse cursor into the start box (centered at the lower part of the screen) to initiate the next part of the story where Jan received one piece of the three context fruit (see Fig. 1.2). The additional visualization with an arrow highlighted that this particular fruit was transferred from the shopping basket to Jan. To further emphasize that Jan actually received one of the three fruit, the transferred fruit was removed from the shopping

Context: Jan's mother was shopping. She bought one banana and two pears. 

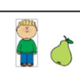
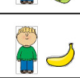
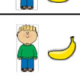
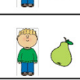
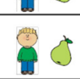
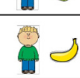
Sentence type	Determiner	Picture	Stimulus sentence	Expected response
false	definite		Of these, Jan received **the** banana.	false
	indefinite		Of these, Jan received **a** pear.	false
felicitous	definite		Of these, Jan received **the** banana.	true
	indefinite		Of these, Jan received **a** pear.	true
infelicitous	definite		Of these, Jan received **the** pear.	false
	indefinite		Of these, Jan received **a** banana.	false

Fig. 2. Examples for the six different conditions resulting from combining the three sentence types (false vs. felicitous vs. infelicitous) with the two determiners (definite vs. indefinite). The color of the word indicating the expected response refers to the color of the respective response box. The words "true" and "false" were not presented on the screen. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

basket. All these measures were taken to make clear that Jan received one of the three initially introduced context fruit, instead of, for example, a novel one.[2] Participants then started their movement into the upper direction and the test sentence appeared below Jan, for example, “Of these, Jan received a pear.” (German original: “Davon hat Jan eine Birne bekommen.”; see Fig. 1.3). This dynamic starting procedure was recommended by Scherbaum and Kieslich (2018), because a static starting condition might lead to less consistent mouse movements. Finally, participants judged the test sentence against the background of the context as “true” or “false” by moving the mouse into the corresponding response box located at the top left and top right corners of the screen (see Fig. 1.4). The boxes were colored in green and red to indicate a “true” and a “false” response. In the example depicted in Fig. 1, a “true” response was appropriate.

Six conditions resulted from crossing the two determiners (definite vs. indefinite) with three sentence types (false vs. felicitous vs. infelicitous; see Fig. 2 for examples of the following). In false sentences, Jan was depicted receiving a different fruit than the one named in the test sentence and a “false” response can be given without processing (anti-)uniqueness of the determiner. In felicitous sentences, the context satisfied (anti-)uniqueness associated with the determiner used in the test sentence. Thus, for definite determiners, the story depicted Jan receiving the unique piece of fruit (the banana), and for indefinite determiners, Jan received one of the non-unique pieces of fruit (a pear). In infelicitous sentences, the context violated the (anti-)uniqueness of the determiner. Thus, for definite determiners, the story depicted Jan receiving one of the non-unique pieces of fruit (a pear), violating the uniqueness presupposition. For indefinite determiners, Jan received the unique piece of fruit (the banana), violating the anti-uniqueness association of the indefinite determiner. The expected response was “false” for both the infelicitous and false sentences, and “true” for the felicitous sentences (see Fig. 2). Stimulus sentences were of the form “Of these, Jan received [determiner: ‘the’ or ‘a’] [fruit].” Seven kinds of fruit were used (banana, lemon, orange, pear, pineapple, plum, and strawberry), all in conjunction with the feminine determiner to avoid the possibility of early disambiguation and to keep the sentences as parallel as possible.

Test sentences started with the word “Davon” (English ”Of these, …“) to emphasize that Jan received a fruit from the three initial fruit of the shopping basket. This facilitates an interpretation of the test sentences and evaluation of the determiner according to the presented local context (i.e., the shopping basket) instead of a more global context (e.g., all available fruit of the world). According to Singh (2011), it is then maximally likely that participants interpret the determiner in the restricted local context where anti-uniqueness and novelty completely coincide.

Because the German word “eine” can also be used as a numeral, participants received error feedback based on the intended pragmatic interpretation of this word in felicitous and infelicitous sentences as an indefinite determiner. (In false sentences, accuracy can be determined on logical grounds and respective error feedback was presented as well.) In Experiment 1, this feedback was provided throughout the experiment; in Experiment 2, it was only provided during a practice phase, but not in the subsequent test phase (for a similar approach to bias interpretations of scalar implicatures, see Tomlinson et al., 2013, and others). It might be true that we induce the indefinite determiner interpretation of the German word “eine” with this feedback. However, this seems necessary to us with respect to our research question, that is, whether we can observe empirical evidence for the psychological reality of the two-step model of processing the indefinite determiner, as suggested by anti-presupposition theory, *if* participants interpret the word “eine” as the indefinite determiner.

Labeling the responses as “true” and “false” was chosen to stay in line with previously published work in the field of scalar implicatures (Bott, Bailey, & Grodner, 2012; Bott & Noveck, 2004; Tomlinson et al., 2013). Note though that these words did never appear on the screen, where the response boxes were simply colored green and red.

1.3.3. Research questions and hypotheses

We were interested in two research questions: (1) Are there processing differences between the definite and the indefinite determiner? and if so (2) What causes these differences? The studies reviewed above suggest that the answer to question (1) is ‘yes.’ As for question (2), anti-presupposition theory claims that the anti-uniqueness interpretation of the indefinite determiner is derived only indirectly, because it relies on a two-step process as described above. By first accessing and then negating the uniqueness-presupposition of the definite determiner, a longer and more complex processing signature for the indefinite determiner is predicted. The overarching question we thus ask is: Is there evidence for the psychological reality of this two-step processing of indefinite determiners? In the next section, we describe how the predictions map onto the measures obtained through mouse-tracking.

The most interesting predictions for our purposes relate to differences between the definite and the indefinite determiner in those conditions where the presupposition-component matters, that is, in the *felicitous* and the *infelicitous* conditions. First, anti-presupposition theory predicts greater processing difficulty for the indefinite relative to the definite determiner, which we predict will lead to an overall larger area under the curve and longer movement times for the indefinite relative to the definite determiner. Second, and most importantly, anti-presupposition theory predicts that the anti-uniqueness associated with the indefinite determiner results from initially considering the uniqueness-presupposition of the definite determiner and then negating it. This, we predict, will lead to larger X_{neg} values for the indefinite relative to the definite determiner. More precisely, consider first the definite determiners (see Fig. 2 for examples). In a felicitous sentence, the uniqueness-presupposition is supported by the context and mouse movements directly head toward the expected “true” response. In an infelicitous sentence, the uniqueness-presupposition is not supported by the context, and mouse movements directly head toward the expected “false” response. In both cases, low X_{neg} values are expected. For the indefinite determiners we assume an initial activation of the uniqueness-presupposition as well. In case of a felicitous sentence, the initial activation of the uniqueness-presupposition is *not* supported by the context, and participants are expected to first show a bias towards a “false” response even when they eventually select the “true” response. In infelicitous sentences, uniqueness is supported by the context, and participants are expected to first show a bias towards a “true” response, even when they eventually select the “false” response. This would be reflected in high X_{neg} values. In sum, anti-presupposition theory predicts larger X_{neg} values for the indefinite compared with the definite determiner for the infelicitous sentences, but surprisingly, also for the felicitous sentences (see also the Online Supplement for a more detailed description).

Predictions are different though for the false control sentences, where the presented picture and the fruit mentioned in the test sentence do not match. Because these sentences can be falsified without

[2] With the description of the context in form of a sentence that introduces exactly one banana with the indefinite determiner, the indefinite determiner in the following test sentence cannot be used to refer to the same banana again. For a further illustration consider the following example:

(a) Last weekend Jan had a view of a new flat.

(b) #A flat was very expensive.

Here, “a” cannot refer to the same flat in (a) and (b). The use of “a” in (a) introduces a flat and with the use of "a" in (b) one expects there to be a second expensive flat. Thus, our procedure worked as much as possible against the introduction of a novel fruit by the participants, in addition to the ones mentioned in the context. Note that this emphasized the infelicity of the indefinite determiner in infelicitous sentences and thus facilitates a “false” response.

processing the (anti-)presupposition, they are expected to cause the least processing problems (compared with the felicitous and infelicitous sentences; see Schwarz, 2016b, for such observation) and no differences are expected between the two determiners in this case.

Further, we expect more difficult processing in the infelicitous than in the felicitous conditions as suggested by results from previous studies (see Burkhardt, 2006; Haviland & Clark, 1974; Schumacher, 2009; van Berkum et al., 1999; but see Clifton, 2013). We predict this to result in a larger area under the curve and longer movement times.

In sum, we expect a larger area under the curve and longer movement times for the indefinite than for the definite determiner in felicitous and infelicitous sentences, but not in false sentences. The respective values are also expected to be smallest for the false sentences, intermediate for felicitous and largest for infelicitous sentences. The theoretically most interesting prediction concerns X_{neg} values: if processing of the indefinite determiner indeed implicates a temporary activation of the uniqueness-presupposition, larger X_{neg} values are expected for the indefinite determiner in the felicitous and infelicitous sentences. This prediction is derived from anti-presupposition theory against the background of the Maximize Presupposition principle, but, importantly, is not made by competing theories of the indefinite determiner (e.g., that the indefinite determiner has its own presupposition; see Kratzer, 2005, and the related discussions in Heim, 1991, 2011).[3] It should further be noted here that two thirds of the trials required a "false" response, and only one third required a "true" response. This unequal distribution might induce a slight bias toward the "false" response box. As a consequence, X_{neg} values would be enlarged for the felicitous sentences (i.e., an overestimation of X_{neg} would be observed), while for infelicitous sentences, this bias reduces the X_{neg} values (which are thus underestimated). While this works against observing processing costs for infelicitous sentences, it is important to note that such a bias does not affect the important differences between definite and indefinite determiners.

2. Experiment 1

In Experiment 1, participants performed the task introduced above. Accuracy feedback was provided in all conditions throughout the experiment to encourage participants to evaluate test sentences based on their relevant pragmatic components, that is, whether the sentences' uniqueness or anti-uniqueness meaning components are met in the context or not.

2.1. Method

2.1.1. Participants

The intended sample size was $n = 20$. Data were collected from twenty-two native speakers of German, of which two participants were excluded because of more than 30% mouse errors (stops or turns) (final sample: mean age = 24.6 years, 15 females, 5 males). All participants reported normal or corrected-to-normal vision and signed written informed consent prior to data collection.

2.1.2. Apparatus and Stimuli

Stimulus presentation and response collection were controlled by a notebook connected to a TFT-screen, with screen resolution set to 1280 × 1024 px. (visible screen size: 53 × 30 cm). Start and response boxes were 60 × 60 px. The start box was centered at the lower part of the screen. The response boxes were located in the upper left/right corner and were colored in green and red to indicate a "true" and a "false" response, but the words "true" and "false" did not appear on the screen. Between the response boxes, the outline of a rectangle was displayed (the shopping basket), and the three context fruits were displayed within it. A picture of Jan was visible below the rectangle. All visual stimuli and sentential materials were presented against a white background with black font color.

2.1.3. Task and Procedure

The participants' task was to evaluate the stimulus sentence against the provided context and the fruit Jan received at the beginning of each trial. Instructions were provided in written form including a graphical illustration of the setup. For this illustration, the pictures of Fig. 1.1.–1.3 were used and additionally described in sentences. The required answers were also illustrated with an infelicitous sentence using a definite determiner to make participants more sensitive to infelicitous uses of determiners without explaining in detail the concept of felicity in the instruction phase. Finally, participants were instructed to perform smooth mouse movements without any stops and backward movements, and they were made familiar with the fruit we used in the experiment and their depiction. Following previous work, participants were instructed to judge the test sentence as "true" or "false" against the visualized context.[4]

Each trial started with the presentation of the initial scene along with its description (see Fig. 1.1). Participants were then required to stay with the mouse cursor in the start box for 500 ms before presentation of the next scene (see Fig. 1.2). From then on, participants needed to start their mouse movement within 5,000 ms to trigger presentation of the test sentence (see Fig. 1.3) and to finish within 12,000 ms. The visual stimulus sentence was presented when the mouse cursor was moved a minimum of 60 px outside the start box into the upper direction, within a corridor of 60 px horizontally centered (to prevent participants from leaving the start box in a diagonal direction). Participants were instructed to select the correct response box as soon as possible then. When participants took longer than 5000 ms to initiate their movement or longer than 12,000 ms to finish the movement, they received the visual feedback "Please answer faster" (German original: "Bitte schneller antworten"). Error feedback ("Wrong answer"; German original: "Falsche Antwort") for undesired responses (see Fig. 2) was also provided visually in all conditions throughout the whole experiment. All feedback was presented centrally in black against the white background for 3000 ms.

For the first three blocks, the green response box was in the left or right corner and the red response box in the respective other corner. The locations were then changed for the remaining three blocks. The initial location was counterbalanced across participants, but each participant performed with both possibilities for two reasons. First, to counteract possible influences of the preferred reading direction on the mouse movements. Second, to counteract an evolving bias toward the red response box (which was required in two thirds of the trials; see above). Participants were informed about the locations prior to each half of the experiment.

Each experimental block comprised 36 trials, resulting from six sentences of each condition (see Fig. 2). As participants performed in six blocks with 36 items each, this results in a total 216 test items. Another 12 un-analyzed practice trials, two of each condition, preceded Blocks 1 and 4. The combination of the context and the test sentence on

[3] Reviewers of this manuscript raised the possibility of other interpretations of the German word "eine" than the intended one as the indefinite determiner. We will consider these interpretations in the General Discussion and detailed accounts are provided in the Online Supplement of this article.

[4] We are aware that there is a difference between logical truth and felicity, but we followed previous work on scalar implicatures (Bott et al., 2004; Tomlinson & Bott, 2013) where the same problem exists. Further, the chosen labels are understood intuitively by the naïve participants, who are usually not trained in pragmatic theory. Explaining the meaning of "felicitous" and "infelicitous" and to ask participants for felicity judgments is difficult, when participants have no/only little linguistic understanding. Note also that this would include telling the participants what felicity means and direct their attention to the research question we were interested in.

each trial was drawn randomly from the pool of all logically possible combinations: With seven types of fruit it is possible to create 84 false combinations with definite and indefinite determiners, and 42 felicitous and infelicitous combinations with definite and indefinite determiners. For the experimental blocks, 36 such combinations were required for each of the six conditions, and each practice block required two additional such combinations per condition. Thus, in sum 40 different combinations were required for each of the six conditions in Experiment 1. Because there is no a priori reason to judge which combinations should be used, we drew randomly from all possible ones (note that increasing the number of items per condition by just one would have required more sentences than were available for felicitous and infelicitous sentences).

The experimenter demonstrated the task prior to the experiment with 12 additional trials using false sentences and emphasized a smooth mouse movement without stops and vertical turns. This was done to reduce errors in mouse movements and thereby to minimize the number of participants that were excluded because of more than 30% mouse movement errors.

Following published work in the field of scalar implicatures (Bott & Noveck, 2004) and negation processing (Dale & Duran, 2011) we did not include filler items. The inclusion of filler items would also imply a reduction of test items, and thus a reduced ability to detect the predicted patterns of the dependent variables due to more noisy data.

2.1.4. Design and analyses

Trials with unspecific errors (initiation time too long, response too slow), stops of the mouse movement (no movement within 200 ms), or backward movements (> 2px) were excluded from the data set first. As mentioned above, two participants were excluded because more than 30% of their trials met the last two criteria. Additional 4.7% of the trials of the remaining 20 participants were excluded because of such errors. Further, trials with incorrect responses were omitted from the data (3.3%). For mouse data analyses, trajectories were centered to a common starting point at $x = 0$ and $y = 120$ (which is the y-coordinate that needed to be exceeded for presentation of the test sentence), with all trajectories ending in the left response box. Calculation of the dependent measures area under the curve, movement time, and X_{neg} was based on raw trajectories, following Kieslich and Henninger (2017). To plot mean trajectories (see Fig. 3), data was time-normalized to 101 timesteps.

Two independent variables were of interest: (1) sentence type (false vs. felicitous vs. infelicitous) and (2) determiner (definite vs. indefinite). The trajectory measures area under the curve, movement time, and X_{neg} were first screened for outliers and trials were excluded if the respective value deviated more than 2.5 standard deviations from the design cell mean (calculated separately for each participant). Error rates were also analyzed to exclude tradeoffs with the time-based measures (see Liesefeld & Janczyk, 2019). Given the hypotheses of the present study, the trajectory measures and error rates were then analyzed with planned contrasts within the 3 × 2 factorial design with sentence type and determiner as repeated-measures. In particular, a set of Helmert contrasts was formulated on sentence type. Contrast 1 compared false sentences with the average of felicitous and infelicitous sentences; Contrast 2 compared felicitous and infelicitous sentences. Using the two orthogonal Helmert contrasts has the advantage to allow for testing interactions of the contrasts as defined on the factor sentence type with the factor determiner in addition. This latter point is important, because we predicted the same increase in X_{neg} values for the indefinite compared with the definite determiner for both the felicitous and the infelicitous sentences. In other words, Contrast 2 should not interact with the factor determiner, in particular for the X_{neg} values. (For completeness, the results of the omnibus 3 × 2 Analyses of Variance are provided in the Online Supplement of this article.) For X_{neg}, one-sample *t*-tests against 0 were additionally calculated (note that by definition X_{neg} cannot become less than 0). When hypotheses allowed

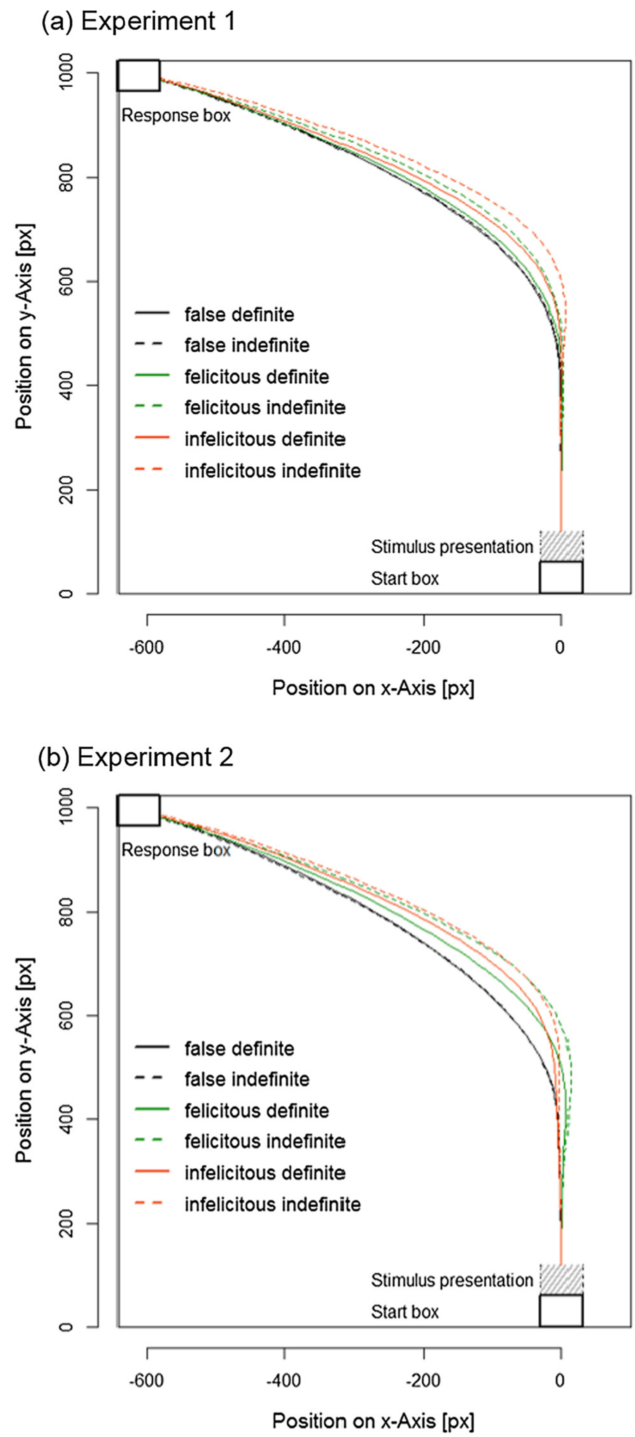


Fig. 3. Mean trajectories for both experiments as a function of sentence type and determiner (for examples of the six different conditions, see Fig. 2). The mouse trajectories visualized in this figure are always mapped to the final response, which was "true" for the felicitous sentences and "false" for the other sentences. This was done to allow an easier comparison of the different trajectories. The stimulus sentence was presented once the mouse cursor vertically left the grey shaded area.

predictions of particular directions (i.e., false < felicitous/infelicitous; felicitous < infelicitous; definite < indefinite determiner), single-tailed *p*-values are reported. Otherwise, tests were two-tailed.

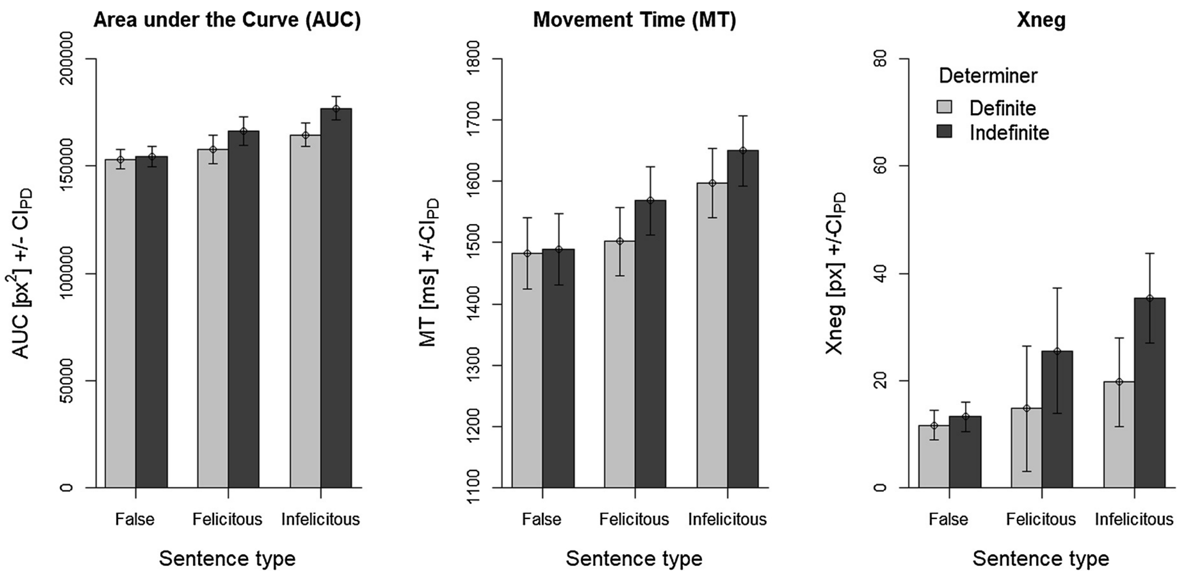


Fig. 4. Means of the trajectory measures as a function of sentence type and determiner from Experiment 1. Error bars are 95% within-subject confidence intervals for the (pairwise) difference between definite and indefinite determiners calculated separately for each dependent measure and sentence type (see Pfister & Janczyk, 2013).

2.2. Results

Fig. 3a visualizes the mean trajectories as a function of sentence type and determiner. For false sentences, no difference between the two determiners is apparent. In contrast, for the felicitous and the infelicitous sentences, the deviation toward the final response location started later for the indefinite determiner, and overall later for infelicitous sentences. In addition, a slight deviation toward the non-target response is apparent for the indefinite determiner.

These impressions are corroborated when considering the trajectory measures (see Fig. 4). Area under the curve (1.71% outliers) was larger for felicitous and infelicitous sentences compared with false sentences, Contrast 1: $t(19) = 5.53$, $p < .001$, $g = 1.24$, $r_{contrast} = 0.79$,[5] and this contrast interacted with the variable determiner, $t(19) = 2.46$, $p = .023$, $g = 0.55$, $r_{contrast} = 0.49$, reflecting a smaller difference between definite and indefinite determiners in false sentences compared with the two other sentence types. The difference between felicitous and infelicitous sentences was also significant, Contrast 2: $t(19) = 3.05$, $p = .003$, $g = 0.68$, $r_{contrast} = 0.57$, but the contrast did not interact with the variable determiner, $t(19) = 1.52$, $p = .145$, $g = 0.34$, $r_{contrast} = 0.33$. Overall, area under the curve was larger for the indefinite than for the definite determiner, $t(19) = 4.02$, $p < .001$, $g = 0.90$, $r_{contrast} = 0.68$. Because Contrast 1 interacted with the variable determiner, the difference between definite and indefinite determiners was assessed for each sentence type. It was not significant for false sentences, $t(19) = 0.57$, $p = .287$, $g = 0.13$, $r_{contrast} = 0.13$, but was so for felicitous, $t(19) = 2.62$, $p = .008$, $g = 0.59$, $r_{contrast} = 0.52$, and infelicitous sentences, $t(19) = 4.56$, $p < .001$, $g = 1.02$, $r_{contrast} = 0.72$.

Movement times (2.34% outliers) were longer for felicitous and infelicitous sentences compared with false sentences, Contrast 1: $t(19) = 5.00$, $p < .001$, $g = 1.12$, $r_{contrast} = 0.75$, but this contrast did not interact with the variable determiner, $t(19) = 1.32$, $p = .203$, $g = 0.29$, $r_{contrast} = 0.29$. Further, the difference between felicitous and infelicitous sentences was significant, Contrast 2: $t(19) = 6.20$, $p < .001$, $g = 1.39$, $r_{contrast} = 0.82$, but this contrast also did not interact with the variable determiner, $t(19) = -0.37$, $p = .714$, $g = -0.08$, $r_{contrast} = 0.08$. Overall, movement times were longer for the indefinite than for the definite determiner, $t(19) = 3.44$, $p = .002$, $g = 0.77$, $r_{contrast} = 0.62$.[6]

X_{neg} values (3.67% outliers) were larger for felicitous and infelicitous sentences compared with false sentences, Contrast 1: $t(19) = 5.61$, $p < .001$, $g = 1.25$, $r_{contrast} = 0.79$, and this contrast interacted with the variable determiner, $t(19) = 3.16$, $p = .005$, $g = 0.71$, $r_{contrast} = 0.59$. The difference between felicitous and infelicitous sentences was also significant, Contrast 2: $t(19) = 2.70$, $p = .007$, $g = 0.60$, $r_{contrast} = 0.53$, but this contrast did not interact with the variable determiner, $t(19) = 0.98$, $p = .341$, $g = 0.22$, $r_{contrast} = 0.22$. Overall, X_{neg} values were larger for the indefinite than for the definite determiner, $t(19) = 3.11$, $p = .003$, $g = 0.69$, $r_{contrast} = 0.58$. Because Contrast 1 interacted with the variable determiner, the difference between definite and indefinite determiners was assessed for each sentence type. It was not significant for false sentences, $t(19) = 1.25$, $p = .114$, $g = 0.28$, $r_{contrast} = 0.28$, but for felicitous sentences, $t(19) = 1.92$, $p = .035$, $g = 0.43$, $r_{contrast} = 0.40$, and for infelicitous sentences, $t(19) = 3.94$, $p < .001$, $g = 0.88$, $r_{contrast} = 0.67$. For all conditions, the X_{neg} values differed significantly from zero, all $t(19)$s ≥ 3.82, all ps $\leq .001$.

Error rates (see Table 1) were larger for felicitous and infelicitous sentences compared with false sentences, Contrast 1: $t(19) = 3.23$, $p = .002$, $g = 0.72$, $r_{contrast} = 0.60$, but this contrast did not interact with the variable determiner, $t(19) = 1.76$, $p = .094$, $g = 0.39$, $r_{contrast} = 0.37$. The difference between felicitous and infelicitous sentences was significant, Contrast 2: $t(19) = 1.96$, $p = .032$, $g = 0.44$, $r_{contrast} = 0.41$, but the contrast did not interacted with the variable determiner, $t(19) = 0.11$, $p = .911$, $g = 0.03$, $r_{contrast} = 0.03$. Overall, error rates were larger for the indefinite than for the definite determiner, $t(19) = 3.62$, $p < .001$, $g = 0.81$, $r_{contrast} = 0.64$.

[5] The effect size $r_{contrast}$ is the partial correlation between dependent variable scores on the individual level and the mean score as predicted by the respective contrast (Rosenthal, Rosnow, & Rubin, 2000). Put simply, $r_{contrast}$ is the contrast analysis-equivalent to η_p.

[6] Although determiner did not interact with the contrasts, we tested the differences between definite and indefinite determiners for each sentence type separately and report them here for completeness. It was not significant for false sentences, $t(19) = 0.26$, $p = .400$, $g = 0.06$, $r_{contrast} = 0.06$, but for felicitous, $t(19) = 2.52$, $p = .001$, $g = 0.56$, $r_{contrast} = 0.50$, and for infelicitous sentences, $t(19) = 1.93$, $p = .034$, $g = 0.43$, $r_{contrast} = 0.40$.

Table 1
Mean error rates (percentage) as a function of sentence type and determiner for both experiments.

	Experiment 1		Experiment 2	
	determiner		determiner	
sentence type	definite	indefinite	definite	indefinite
false	0.61	2.29	1.48	1.03
felicitous	1.29	4.68	2.62	5.75
infelicitous	3.73	7.36	3.75	3.16

2.3. Discussion

Experiment 1 revealed several interesting results. First, area under the curve was larger and movement times were longer for indefinite than for definite determiners in felicitous and infelicitous sentences. Thus, there appears to be an increased processing effort associated with indefinite determiners. In addition, and most importantly, X_{neg} values were reliably larger for indefinite than for definite determiners, and this was true for both felicitous and infelicitous sentences. This suggests that the uniqueness-presupposition of the definite determiner is at least briefly and temporarily activated and evaluated by participants upon encountering an indefinite determiner, as is suggested by anti-presupposition theory.

Second, area under the curve was smaller and movement times were shorter in the false compared to the felicitous and infelicitous conditions, and no difference between determiners was observed in the false condition. Thus, mere falsification of a sentence, which does not require assessing a presupposition-related meaning component, comes with the least processing difficulties.

Third, overall area under the curve was smaller and movement times were shorter in the felicitous condition compared to the infelicitous condition. Thus, it appears that processing is more difficult in case of infelicitous compared with felicitous sentences (see also Hertrich et al., 2015; Kirsten et al., 2014; Tiemann et al., 2011; and others).

In sum, these results are well in line with our predictions. However, the error feedback provided throughout the whole experiment adds some oddness to the experiment, because in natural language use, a speaker is often not provided immediate feedback concerning her interpretation of an utterance. In Experiment 2, we borrowed an approach used by Tomlinson et al. (2013, Exp. 1) and "primed" participants toward the intended indefinite interpretation of the German word "eine" only during the initial practice phase. Thus, the practice phase was extended and feedback was provided in all conditions in this practice phase, but not during the subsequent test blocks. Since this may reduce the impact of the suspected processing difficulties on the mouse trajectories, we increased sample size as a countermeasure.

3. Experiment 2

This experiment largely resembled Experiment 1, but error feedback was only provided during the practice phase, but not during the subsequent test blocks.

3.1. Method

3.1.1. Participants

The intended sample size in this experiment was increased to $n = 50$. Data were collected from fifty-three native speakers of German, of which three participants were excluded because of more than 30% mouse errors (final sample: mean age = 24.2 years, 39 females, 11 males). All participants reported normal or corrected-to-normal vision and signed written informed consent prior to data collection.

3.1.2. Stimuli, Apparatus, Task, Procedure, Design, and Analyses

The experiment resembled Experiment 1 with the following changes: Participants received error feedback only during the practice trials prior to the experimental blocks. The practice phase before Block 1 was increased to 48 trials made of 24 different trials repeated twice in a random order. Thus, we required additional six combinations (of context and test sentence) per condition in this experiment. For the experimental blocks, 36 combinations of context and test sentence were required for each of the six conditions. In sum, Experiment 2 required all available 42 different combinations for felicitous and infelicitous sentences, which were presented in a random order. The 42 different false combinations for each determiner were randomly drawn from the available false combinations (see Experiment 1 for more details). We removed 8.2% of the trials because of unspecific errors or mouse movement errors.

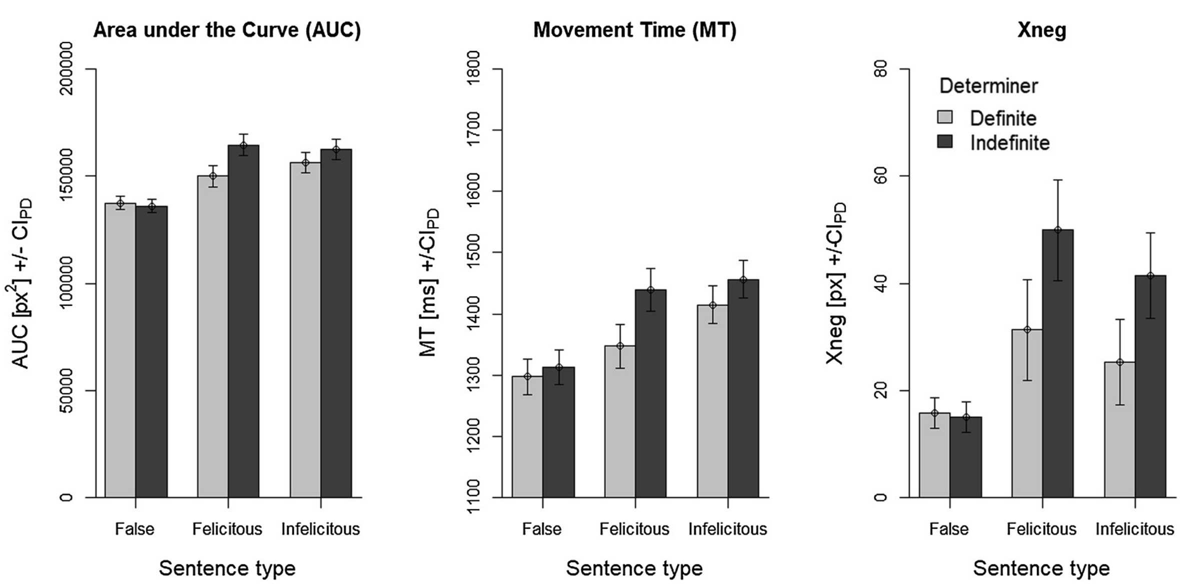


Fig. 5. Means of the trajectory measures as a function of sentence type and determiner from Experiment 2. Error bars are 95% within-subject confidence intervals for the (pairwise) difference between definite and indefinite determiners calculated separately for each dependent measure and sentence type (see Pfister & Janczyk, 2013).

3.2. Results

Mean trajectories are visualized in Fig. 3b and means of all trajectory measures are visualized in Fig. 5. The general impression is very similar to the results of Experiment 1. All dependent measures showed higher values for felicitous and infelicitous sentences compared with false sentences, and higher values for indefinite than for definite determiners, but only within the felicitous and infelicitous sentences. A slight deviation toward the non-target response for indefinite determiners seems also to be present.

The area under the curve (1.62% outliers) was larger for felicitous and infelicitous sentences compared with false sentences, Contrast 1: $t(49) = 9.60$, $p < .001$, $g = 1.36$, $r_{contrast} = 0.81$, and this contrast also interacted with the variable determiner, $t(49) = 4.73$, $p < .001$, $g = 0.67$, $r_{contrast} = 0.56$. The difference between felicitous and infelicitous sentences was not significant, Contrast 2: $t(49) = 0.56$, $p = .286$, $g = 0.08$, $r_{contrast} = 0.08$, but the contrast interacted with the variable determiner, $t(49) = -2.65$, $p = .012$, $g = -0.37$, $r_{contrast} = -0.35$, reflecting a larger difference between the definite and the indefinite determiner in the felicitous compared with the infelicitous condition. Overall, the area under the curve was larger for the indefinite than for the definite determiner, $t(49) = 5.85$, $p < .001$, $g = 0.69$, $r_{contrast} = 0.57$. Because the contrasts interacted with the variable determiner, the difference between definite and indefinite determiners was assessed for each sentence type. It was not significant for false sentences, $t(49) = -0.88$, $p = .190$, $g = -0.13$, $r_{contrast} = 0.13$, but was so for felicitous, $t(49) = 5.91$, $p < .001$, $g = 0.84$, $r_{contrast} = 0.64$, and infelicitous sentences, $t(49) = 2.50$, $p = .008$, $g = 0.35$, $r_{contrast} = 0.34$.

Movement times (2.22% outliers) were longer for felicitous and infelicitous sentences compared with false sentences, Contrast 1: $t(49) = 7.61$, $p < .001$, $g = 1.08$, $r_{contrast} = 0.74$, and this contrast interacted with the variable determiner, $t(49) = 3.43$, $p = .001$, $g = 0.49$, $r_{contrast} = 0.44$. Further, the difference between felicitous and infelicitous sentences was significant, Contrast 2: $t(49) = 3.78$, $p < .001$, $g = 0.54$, $r_{contrast} = 0.48$, but this contrast did not interact with the variable determiner, $t(49) = -1.86$, $p = .069$, $g = -0.26$, $r_{contrast} = 0.26$. Overall, movement times were longer for the indefinite than for the definite determiner, $t(49) = 5.77$, $p < .001$, $g = 0.82$, $r_{contrast} = 0.64$. Because Contrast 1 interacted with the variable determiner, the difference between definite and indefinite determiners was assessed for each sentence type. It was not significant for false sentences, $t(49) = 1.06$, $p = .146$, $g = 0.15$, $r_{contrast} = 0.15$, but it was significant for felicitous, $t(49) = 5.21$, $p < .001$, $g = 0.74$, $r_{contrast} = 0.60$, and infelicitous sentences, $t(49) = 2.75$, $p = .004$, $g = 0.39$, $r_{contrast} = 0.37$.

X_{neg} values (3.92% outliers) were larger for felicitous and infelicitous sentences compared with false sentences, Contrast 1: $t(49) = 8.82$, $p < .001$, $g = 1.25$, $r_{contrast} = 0.78$ and this contrast interacted with the variable determiner, $t(49) = 4.62$, $p < .001$, $g = 0.65$, $r_{contrast} = 0.55$. X_{neg} values were not significantly larger for infelicitous compared to felicitous sentences, Contrast 2: $t(49) = -1.58$, $p = .940$ $g = -0.22$, $r_{contrast} = 0.22$, and this contrast's interaction with the variable determiner was not significant, $t(49) = -0.49$, $p = .312$, $g = -0.07$, $r_{contrast} = 0.07$. Overall, X_{neg} values were larger for the indefinite than for the definite determiner, $t(49) = 4.86$, $p < .001$, $g = 0.69$, $r_{contrast} = 0.57$. Because Contrast 1 interacted with the variable determiner, the difference between definite and indefinite determiners was assessed for each sentence type. It was not significant for false sentences, $t(49) = -0.58$, $p = .282$, $g = -0.08$, $r_{contrast} = 0.08$, but it was significant for felicitous sentences, $t(49) = 3.96$, $p < .001$, $g = 0.56$, $r_{contrast} = 0.49$, and for infelicitous sentences, $t(49) = 4.05$, $p < .001$, $g = 0.57$, $r_{contrast} = 0.50$. For all conditions, the X_{neg} values differed significantly from zero, all $t(49)\text{s} \geq 6.61$, all $ps \leq .001$.

Error rates (see Table 1) were larger for felicitous and infelicitous sentences compared with false sentences, Contrast 1: $t(49) = 2.99$, $p = .002$, $g = 0.42$, $r_{contrast} = 0.39$, and this contrast interacted with the variable determiner, $t(49) = 3.23$, $p = .002$, $g = 0.46$, $r_{contrast} = 0.42$. The difference between felicitous and infelicitous sentences was not significant, Contrast 2: $t(49) = -1.30$, $p = .100$, $g = -0.18$, $r_{contrast} = 0.18$, and the contrast did not interact with the variable determiner, $t(49) = -1.17$, $p = .247$, $g = -0.17$, $r_{contrast} = 0.17$. Overall, error rates were larger for the indefinite than for the definite determiner, $t(49) = 2.08$, $p = .043$, $g = 0.29$, $r_{contrast} = 0.28$. Because Contrast 1 interacted with the variable determiner, the difference between definite and indefinite determiners was assessed for each sentence type. It was not significant for false, $t(49) = -1.35$, $p = .092$, $g = -0.19$, $r_{contrast} = 0.19$, and infelicitous sentences, $t(49) = -0.36$, $p = .361$, $g = -0.05$, $r_{contrast} = 0.05$, but was so for felicitous sentences, $t(49) = 1.90$, $p = .032$, $g = 0.27$, $r_{contrast} = 0.26$.

3.3. Discussion

Providing feedback only in the practice block sufficed to make the participants take the context into account and to interpret the determiners as intended, and a largely similar pattern of results as observed in Experiment 1 was revealed in Experiment 2. We again observed a larger area under the curve and longer movement times for the indefinite compared with the definite determiner within the felicitous and infelicitous conditions. Importantly, X_{neg} values were higher for the indefinite than for the definite determiner in both the felicitous and the infelicitous conditions. This is again consistent with the claim that the uniqueness-presupposition of the definite determiner receives a brief initial activation upon encountering the indefinite determiner.

The results are less clear than those of Experiment 1 with regard to more difficult processing in infelicitous compared with felicitous sentences (see also Clifton, 2013). Area under the curve was smaller and movement times were shorter in the false condition than in the felicitous and infelicitous conditions, and no difference between determiners was observed for the false condition. Movement times were also longer for the infelicitous compared with the felicitous conditions, thus suggesting more processing difficulties in the former compared with the latter condition, but the corresponding contrast for area under the curve was not significant (though descriptively in the predicted direction).

4. General discussion

The present study provides evidence for the psychological reality of a formal theory of presupposition generation, known as Maximize Presupposition. To this end, we used mouse-tracking, a method that has not been applied to this phenomenon so far, but offers the possibility to analyze general processing difficulties as well as an indirect measure of the interpretations entertained by a reader in the course of sentence processing. Participants were asked to evaluate sentences with definite and indefinite determiners in three types of test sentences (see also Fig. 2): (1) literally false sentences, (2) literally true and pragmatically felicitous sentences, and (3) literally true, but pragmatically infelicitous sentences, because of a violation of their presuppositional meaning component (uniqueness or anti-uniqueness). Across both experiments, the main result is that processing of the indefinite determiner is more difficult than processing the definite determiner, and involves a temporary initial activation of the uniqueness-presupposition. This latter result can be seen as support for anti-presupposition theory.

4.1. Summary of the results and their theoretical implications

Several results were clear-cut across both experiments. As expected, false sentences came with the least processing difficulties compared with the felicitous and infelicitous sentences. Further, except for the false sentences, the indefinite determiner was associated with more difficult processing as reflected in longer movement times and a larger area under the curve. Finally, and most important for the present purposes, X_{neg} values were larger for the indefinite than for the definite determiner in both felicitous and infelicitous sentences.

The X_{neg} result is particularly informative, because it provides evidence that processing the indefinite determiner actually involves an initial activation of the uniqueness-presupposition of the definite determiner. This prediction is derived uniquely from anti-presupposition theory and the Maximize Presupposition principle (Heim, 1991). According to this principle, indefinite determiners carry their anti-uniqueness meaning component as an anti-presupposition, which is derived indirectly in two steps: First, the uniqueness-presupposition of the definite determiner is considered, and second, it is then negated. This predicts that participants are initially more biased toward the non-target response, but subsequently correct their decision to the final response. This would be reflected in larger X_{neg} values for the indefinite compared with the definite determiner in both the felicitous and infelicitous sentences (see the Online Supplement for a more detailed description), and is exactly what we observed in both our experiments. This result is particular noteworthy for the infelicitous sentences, where the initial bias was observed, even though the context facilitates a "false" response. To avoid misunderstandings, this interpretation is not meant to imply that participants evaluate the initial uniqueness-presupposition during the sentence comprehension process consciously.

Considering the rather small effects observed, we suggest that the uniqueness-presupposition receives some initial, but temporarily restricted activation. Still, this activation was measurable from the mouse movements, and left a trace on the participants' behavior. In contrast, the initial biases seem much larger for implicatures (see Tomlinson et al., 2013). This may point to a larger effort required to enrich the initial meaning with the additional meaning components in the case of scalar implicatures. Although it should again be noted that the two suggested steps in processing scalar implicatures and the indefinite determiner are different, future studies may directly compare both pragmatic phenomena and reduce methodological differences between the studies. For example, we used a dynamic procedure for the stimulus presentation (see Scherbaum & Kieslich, 2018, for advantages of this), and participants were required to leave the start box with a movement directing straight-up. Thus, the stimulus was perceived and read during the movement itself, and the required straight initial movement may have reduced the opportunity for large deviations.

Somewhat inconsistent are the results concerning the processing difficulties for infelicitous compared with felicitous sentences. In Experiment 1, infelicitous sentences had longer movement times and a larger area under the curve than felicitous sentences, thus suggesting greater processing difficulties. In Experiment 2, however, this was only observed in movement times, but not for area under the curve. Remember that false and infelicitous sentences both required a "false" response. Consequently, two factors may have increased processing difficulties of infelicitous sentences: First, the unnaturalness of classifying an infelicitous sentence with the "false" response, and second, the ambiguity arising from two sentence types requiring the same response (compared with the felicitous sentences, which were the only ones requiring a "true" response). Thus, it is conceivable that the variables (movement times, area under the curve) indicating larger processing difficulties for the infelicitous sentences were artificially increased to some degree, although the bias toward a "false" response (which was required in two thirds of the trials; see Introduction) might simultaneously have increased values for the felicitous sentences. In sum, however, the most warranted conclusion would be that processing difficulties were similar in felicitous and in infelicitous sentences, contrary to what has been observed in some other studies (e.g., Hertrich et al., 2015; Kirsten et al., 2014; and others). A similar observation, that is, no processing disadvantage for infelicitous conditions, was reported in Experiments 1 and 3 of Clifton (2013), who argued that this difference is only observed with sufficient cognitive workload (as induced in his Experiments 2 and 4 with the addition of a mental arithmetic task).

4.2. Objections and possible alternative accounts

In this section we discuss two concerns that may affect the interpretation of the results: first, that the error feedback provided during the test trials in Experiment 1 and the practice trials in Experiment 2 coerced participants into an indefinite interpretation of the German word "eine" (English 'a') in a way that is not ecologically valid, and second, that the experimental setup did not successfully exclude alternative interpretations of this word.

In the practice phases of both experiments and throughout Experiment 1, participants received error feedback in case they responded in an undesired way (see Fig. 2). This was done to ensure that participants actually entertain the desired interpretation of the word "eine" as an indefinite determiner, thereby enhancing the internal validity of the experiment. A similar way has been taken by previous research to coerce participants into the particular behavior of current interest (Bott & Noveck, 2004; Bott et al., 2012; Tomlinson et al., 2013), and thus seems to be a reasonable way to achieve the desired interpretation. It should also be noted that the critical results remained the same in Experiment 2, even when error feedback was removed from all test trials and only provided during the 48 practice trials.

The most important point, however, is that the aim of this study was not to investigate all the possible interpretations of the German word "eine", but rather to examine the processing of this word on its indefinite reading. That is, in contexts that induce the desired indefinite reading, how is this word processed, as opposed to the definite determiner? In other words, we did not investigate "Under which circumstances interpret people the word 'eine' as an indefinite determiner, as opposed to, say, a number word?", although this is certainly an interesting question as well.

A further objection in relation to the error feedback is that participants may only have learned how to respond in the course of the experiment. Under the assumption that encountering the word "eine" causes uncertainty, initial deviations toward the wrong response may occur on a subset of trials. This could explain the larger X_{neg} values in the indefinite determiner conditions within the felicitous and infelicitous sentences. However, this reasoning suggests that mean trajectories for indefinite determiner conditions are a mixture of two types of trajectories: (1) those directly heading toward the correct response and (2) those first heading toward the non-target response. This would be reflected in a bimodal distribution of, for example, area under the curve. However, area under the curve did not show signs of bimodality in both experiments, as all bimodality coefficients were ≤ 0.40 and thus point to a unimodal distribution (Freeman & Dale, 2013; Pfister, Schwarz, Janczyk, Dale, & Freeman, 2013). Thus, at present we do not consider this alternative explanation likely, but future research should certainly pick up this issue.

Another and equally important question is: Are there alternative interpretations of the word "eine" that make the same predictions as those we made from our assumptions based on anti-presupposition theory? We will discuss three possible interpretations in the following: (1) a literal interpretation, (2) an interpretation as a numeral (e.g., "one banana"), and (3) an interpretation as a novel object that was not introduced in the given context (e.g., "some or other banana"). For further details and a discussion of the definite determiner in these cases, we refer the reader to the Online Supplement (Scenarios 1–3). Not taking the exact number into account, a *"literal" interpretation* of the word "eine" (the at-issue reading without anti-uniqueness or other enrichments) is compatible with the picture in both felicitous and infelicitous sentences. If we therefore assume, similar to the first step in two-step models of scalar implicatures' processing (see Bott & Noveck, 2004; Tomlinson et al., 2013), that the first step of processing consists of entertaining only the literal reading, an initial response tendency towards the "true" response may result. If then participants want to avoid response errors, they need to reverse their movement direction toward the "false" response in an infelicitous sentence. Thus, X_{neg}

values are expected to be high in infelicitous sentences, but low in felicitous sentences, contrary to what we observed. The same reasoning applies if participants endorse an interpretation of the German word "eine" as a *numeral* ("one"), because Jan always received exactly one banana or pear. Finally, it is possible that participants interpreted the mentioned fruit still as a *novel* one, that is, one not introduced in the context already (see Grønn & Sæbø, 2012; Heim, 1983), although we set up the context (including the descriptive sentences, the pictorial illustration, and starting the test sentences with "Davon…." [engl. "Of these, …"]) in a way to prevent this. If participants then interpret the sentence as "Jan received some or other banana/pear", the initial response direction would be "true". However, then the same reasoning as for the previous two interpretations applies and the same prediction results for the X_{neg} values.

In sum, these three alternative readings predict small X_{neg} values in felicitous sentences and large X_{neg} values in infelicitous sentences when encountering and interpreting the German word "eine" in these ways. Only the assumption of an initial activation of the uniqueness-presupposition as derived from anti-presupposition theory predicts larger X_{neg} values for the indefinite than for the definite determiner in both felicitous and infelicitous sentences—and this is what we observed. Depending on how the definite determiner is processed in the case of the three alternatives discussed above, further diverging predictions may arise (in particular when one assumes that only existence is checked against the current context). For more details, please consider the Online Supplement.

4.3. Limitations and future extensions

Although we are convinced that the present study provides important insights into presupposition processing, there are several potential limitations and possible extensions for future work.

One difference between the German definite and indefinite feminine determiner is their length, with the indefinite determiner comprising more letters and syllables. Thus, one may object that the increased processing difficulty results from longer reading and/or comprehension time of the indefinite determiner. This is, however, unlikely for at least two reasons. First, no difference between the determiners was measured for the false sentences, which served as a control condition. Second, such account would predict differences only in time-based measures, but the theoretically most important differences were observed in X_{neg} values.

Yet, one may object that the results from the false sentences are not conclusive, because the strategy of just noting that the last word mismatches the picture is sufficient to judge the sentences as false. While theoretically this is a possible strategy, it is inefficient within the context of the present experiment, because it can be only applied in one third of the trials in both experiments, and participants never knew in advance when they can and when they cannot successfully employ it. Further, consider what follows if participants indeed always employed this strategy. In the felicitous and the infelicitous sentences, the initial response would then be the "true" response. This is already the correct response for felicitous sentences. To avoid committing errors in the infelicitous sentences, a reversal of the movement direction toward the "false" response is then required (note that to distinguish felicitous and infelicitous sentences it is also necessary to consider the whole context). Thus, for felicitous sentences, one would predict small X_{neg} values (for both types of determiners), while for infelicitous sentences large X_{neg} values are expected (for both types of determiners). This differs from our predictions and also from the observed results. Thus, we are confident that participants did not follow such a strategy (for more details, see Scenario 4 in the Online Supplement).

Further limitations of the present study result from the choice of the language and the particular presupposition trigger we focused on. More precisely, we investigated only the definite and the indefinite determiner, and this only in the German language. One reason for this is that the determiner, and in particular the indefinite one, is an interesting case from a linguistic perspective and its status is unclear with regard to whether it triggers a presupposition or can better be viewed as an anti-presupposition. As such, the Maximize Presupposition principle is most prominently applied to the case of determiners. However, it also applies to other presupposition triggers. For example, "both" has a competitor, namely "all" and Maximize Presupposition accounts for the fact that the possibility of using a sentence with "both" blocks the possibility of using a sentence with "all". It is as yet unclear whether in these cases a two-step model also applies, but such additional triggers are certainly a vital topic for future research. Further, future studies should be carried out in different languages than German, to explore the generality of our results across different languages.

4.4. Conclusion

The Maximize Presupposition principle and the consequent anti-presupposition theory for the indefinite determiner suggest that the anti-uniqueness of this determiner results only indirectly via a two-step processing: An initial consideration of the uniqueness-presupposition of the definite determiner followed by its subsequent negation. The two experiments of our study provide experimental evidence for the psychological reality of this assumption. Thus, the data support this principle and add an important piece of evidence to the experimental pragmatics literature. They also support the idea that presuppositions and anti-presuppositions differ and must be treated as different pragmatic phenomena with systematic differences in their processing signature.

5. Author note

This research was supported by the German Research Foundation (DFG) via Project B2 of the SFB 833 "The constitution of meaning".

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.104024.

References

Alonso-Ovalle, L., Menéndez-Benito, P., & Schwarz, F. (2011). Maximize presupposition and two types of definite competitors. *Proceedings of North East Linguistic Society Conference (NELS), 39*, 29–40.
Altmann, G., & Steedman, M. (1988). Interaction with context during sentence processing. *Cognition, 30*, 191–238.
Beaver, D., & Geurts, B. (2012). Presupposition. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.). *Semantics: An international handbook of natural language meaning* (pp. 2432–2459). Berlin: Mouton de Gruyter.
Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language, 66*, 123–142.
Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*, 437–457.
Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language, 98*, 159–168.
Chemla, E. (2009). Presuppositions of quantified sentences: Experimental data. *Natural Language Semantics, 17*, 299–340.
Clifton, C. J. (2013). Situational context affects definiteness preferences: Accommodation of presuppositions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 487–501.
Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science, 35*, 983–996.
Freeman, J. B., & Ambady, N. (2010). Mousetracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*, 226–241.
Freeman, J. B., & Ambady, N. (2011). When two become one: Temporally dynamic integration of the face and voice. *Journal of Experimental Social Psychology, 47*, 259–263.
Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods, 45*, 83–97.
Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology, 2*, 59.
Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.). *Speech acts*

[syntax and semantics 3] (pp. 41–58). New York, NY: Academic Press.
Grønn, A., & Sæbø, K. J. (2012). A, the, another: A game of same and different. *Journal of Logic, Language and Information, 21*, 75–95.
Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13*, 512–521.
Hawkins, J. A. (1978). *Definiteness and indefiniteness.* London, UK: Routledge.
Heim, I. (1983). On the projection problem for presuppositions. In D. P. Flickinger (Ed.). *Proceedings of WCCFL 2* (pp. 114–125). Stanford University, CA: CSLI Publications.
Heim, I. (2011). Definiteness and indefiniteness. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.). *Semantics: An international handbook of natural language meaning* (pp. 996–1025). Berlin: De Gruyter Mouton.
Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar.* New York, NY: Wiley-Blackwell.
Heim, I. (1991). Artikel und Definitheit. In A.v. Stechow, & D. Wunderlich (Eds.). *Semantik/Semantics. Ein internationales Handbuch der zeitgenössischen Forschung. An international handbook of contemporary research* (pp. 487–535). Berlin: De Gruyter Mouton.
Hertrich, I., Kirsten, M., Tiemann, S., Beck, S., Wühle, A., Ackermann, H., et al. (2015). Context dependent impact of presuppositions on early magnetic brain responses during speech perception. *Brain and Language, 149*, 1–12.
Huang, Y., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes, 26*, 1161–1172.
Janczyk, M., Pfister, R., & Kunde, W. (2013). Mice move smoothly: Irrelevant object variation affects perception, but not computer-mouse actions. *Experimental Brain Research, 231*, 97–106.
Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods, 49*, 1652–1667.
Kirsten, M., Tiemann, S., Seibold, V. C., Hertrich, I., Beck, S., & Rolke, B. (2014). When the polar bear encounters many polar bears: Event-related potential context effects evoked by uniqueness failure. *Language, Cognition and Neuroscience, 29*, 1147–1162.
Kratzer, A. (2005). Building resultatives. In C. Maienborn, & A. Wöllstein-Leisten (Eds.). *Event arguments in syntax, semantics, and discourse* (pp. 177–212). Tübingen, Germany: Niemeyer.
Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy tradeoffs (?). *Behavior Research Methods, 51*, 40–60.
Murphy, G. (1984). Establishing and accessing referents in discourse. *Memory Cognition, 12*, 489–497.
Percus, O. (2006). Antipresuppositions. In A. Ueyama (Ed.). *Theoretical and empirical studies of reference and anaphora: Toward the establishment of generative grammar as an empirical science* (pp. 52–73). Japan Society for the Promotion of Science.
Pfister, R., & Janczyk, M. (2013). Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. *Advances in Cognitive Psychology, 9*, 74–80.
Pfister, R., Janczyk, M., Wirth, R., Dignath, D., & Kunde, W. (2014). Thinking with portals: Revisiting kinematic cues to intention. *Cognition, 133*, 464–473.
Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. B. (2013). Good things peak in pairs: A note on the bimodality coefficient. *Frontiers in Quantitative Psychology and Measurement, 4*, 700.
Roettger, T. B., & Stoeber, M. (2017). Manual response dynamics reflect rapid integration of intonational information during reference resolution. In G. Gunzelmann, (Ed.). *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 3010–3015). Austin, Texas: Cognitive Science Society.
Roettger, T. B., & Franke, M. (2018). Dynamic speech adaptation to unreliable cues during intonational processing. In C. Kalish, (Ed.). *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 966–971). Austin, Texas: Cognitive Science Society.
Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach.* Cambridge: University Press.
Sauerland, U. (2008). Implicated presuppositions. In A. Steube (Ed.). *The discourse potential of underspecified structures* (pp. 581–600). Berlin: Mouton de Gruyter.
Sauerland, U., Tamura, A., Koizumi, M., & Tomlinson, J. M. (2015). Tracking down disjunction. In D. Bekki (Ed.). *Preproceedings of logic and engineering of natural language semantics 12*. Tokyo: Ochanomizu University.
Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition, 115*, 407–416.
Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods, 50*, 2097–2110.
Schlenker, P. (2012). Maximize presupposition and Gricean reasoning. *Natural Language Semantics, 20*, 391–429.
Schumacher, P. B. (2009). Definiteness marking shows late effects during discourse processing: Evidence from ERPs. *Lecture Notes in Computer Science, 5847*, 91–106.
Schwarz, F. (2007). Processing presupposed content. *Journal of Semantics, 24*, 373–416.
Schwarz, F. (2016b). False but slow: Evaluating statements with non-referring definites. *Journal of Semantics, 33*, 177–214.
Schwarz, F. (2016a). Experimental work in presupposition and presupposition projection. *Annual Review of Linguistics, 2*, 273–292.
Singh, R. (2011). Maximize Presupposition! and local contexts. *Natural Language Semantics, 19*, 149–168.
Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences, 13*, 360–366.
Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academic of Science, 102*, 10393–10398.
Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic, 2*, 447–457.
Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., et al. (2011). Psycholinguistic evidence for presuppositions: On-line and off-line data. *Proceedings of Sinn & Bedeutung, 15*, 581–595.
Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language, 69*, 18–35.
Tomlinson, J. M., & Bott, L. (2013). How intonation constrains pragmatic inference. *Proceedings Cognitive Science Society, 35*, 3569–3575.
van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147–182.
van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*, 235–248.
Wirth, R., Pfister, R., Janczyk, M., & Kunde, W. (2015). Through the portal: Effect anticipation in the central bottleneck. *Acta Psychologica, 160*, 141–151.

10.2 Study 2

Manuscript accepted for publication.

Schneider, C., & Janczyk, M. (accepted). Capacity limitations of processing presuppositions triggered by determiners. *Acta Psychologica.*

Capacity limitations of processing presuppositions triggered by determiners

[a]Cosima Schneider (ORCID: 0000-0002-4314-0885)

&

[b]Markus Janczyk (ORCID: 0000-0002-9958-3220)

[a]Department of Psychology, University of Tübingen, Tübingen, Germany
[b]Department of Psychology, University of Bremen, Bremen, Germany

Author note: This research was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) via Project B2 of the Collaborative Research Center 833 "The constitution of meaning". We thank two anonymous reviewers for their helpful and constructive comments on an earlier version of this manuscript.

Corresponding Author:
Cosima Schneider
Department of Psychology
University of Tübingen
Schleichstraße 4
72076 Tübingen
Germany
Phone: +49 (0)7071 29 78359
Email: cosima.schneider@uni-tuebingen.de

Abstract

Definite determiners trigger existence- and uniqueness-presuppositions, that is, the speaker assumes that it is taken for granted that there exists exactly one of the mentioned objects in the relevant discourse. Indefinite determiners are associated with anti-uniqueness, that is, that there are several of the mentioned object. Applying the Maximize Presupposition principle, this additional meaning component arises as an anti-presupposition and involves first considering the definite determiner's uniqueness-presupposition and then its negation. We here investigate processing of the two determiners in more detail and ask whether this processing is automatic or requires limited central capacities. To do so, we employed the Psychological Refractory Period (PRP) approach and the locus of slack-logic. We observed more difficult processing for the indefinite compared to the definite determiner in felicitous sentences, and also in infelicitous compared to felicitous sentences. Further, immediate processing of presuppositions appears capacity-limited. These results support the Maximize Presupposition principle and are an important step forward towards understanding cognitive processing of presuppositions.

Key words: presuppositions ; experimental pragmatics ; dual-task ; PRP

Capacity limitations of processing presuppositions

There are three bottles and one glass of water on the table, and somebody says ‘Give me the bottle!’ – Although this is a grammatically sound sentence and you are familiar with the semantics of all words, something is odd. This oddness results from you expecting that using the definite determiner “the” refers to an object which is unique in the current context of discourse (in addition to the fact that a bottle exists at all). As a consequence, while not logically false, the example sentence is *infelicitous*. In contrast, ‘Give me the glass!’ would be *felicitous*. Technically, the assumption of existence and uniqueness associated with the definite determiner are called *presuppositions* (Heim, 1991, 2011) and the definite determiner “the” acts as their *trigger*. While the presuppositions of the definite determiner are largely agreed on, there is much more discussion about why the indefinite determiner, in contrast, is associated with anti-uniqueness (e.g., Alonso-Ovalle, Menéndez-Benito, & Schwarz, 2011).

Presuppositions of determiners. In general, presuppositions convey information that is already assumed or taken for granted by the discourse participant. In Example (1) the speaker assumes that the listener takes it for granted that Anna has scored before yesterday. This additional meaning component is triggered by the word “again”.

(1) Yesterday, Anna scored again.

Frege (1892) was the first to discuss presuppositions in connection with definite descriptions, and since then, the literature suggests that the definite determiner triggers existence- and uniqueness-presuppositions (e.g., Heim, 1991, 2011). When the presuppositions are met by the context, the definite determiner must be used as in Example (2), otherwise it becomes odd as in Example (3), where the use of the indefinite determiner is infelicitous.

(2) The father of the soccer coach laughed.

(3) # A father of the soccer coach laughed.

The inappropriateness of the indefinite determiner in Example (3) is often accounted for by the *Maximize Presupposition* principle (Heim, 1991): *Presuppose as much as possible!* (see also

Chemla, 2009; Sauerland, 2008; Schlenker, 2012). This principle – an addition to the Gricean maxims of conversation (Grice, 1975) – requires speakers to always use the felicitous sentence with the strongest presupposition among a set of alternatives, if the speaker knows that these presuppositions are fulfilled. The definite determiner is presuppositionally stronger because it triggers two presuppositions (existence and uniqueness). In contrast, the indefinite determiner is associated with anti-uniqueness. This additional meaning component is not a presupposition, but instead arises from pragmatic reasoning over alternatives. In this way, Maximize Presupposition accounts for the fact that indefinite determiners of the form "a X" (where X is a noun phrase) are infelicitous when it is common ground[1] that there is exactly "one X". In Example (3), the speaker chose the indefinite determiner. Because the indefinite determiner does not presuppose anything it is presuppositionally weaker than its stronger alternative, the definite determiner. By choosing a weaker item from two alternatives that can be ordered on a scale (similar to scalar implicatures), it is implied that the presupposition of the stronger alternative (the definite determiner) does not hold. In Example (3) this means that the coach does not have a unique father. The oddness of the utterance follows from the conflict with our common knowledge. Accordingly, as an "anti-presupposition" (Percus, 2006), the indefinite determiner leads to anti-uniqueness only indirectly in two steps: (1) evaluation of the uniqueness-presupposition of the definite determiner and (2) the subsequent negation of this presupposition. The second step leads to more difficult processing compared with the definite determiner.

Previous experimental work. Previous studies provided evidence for an immediate processing of presuppositions starting already on the trigger. For example, Schwarz (2007) employed self-paced reading to compare processing of sentences including the word *auch* (too)

[1] The common ground is defined as the set of propositions believed to be true by all participants of a conversation. More formally speaking, a sentence *p* presupposes *q* if the use of *p* is inappropriate when *q* is not in the common ground (Stalnaker, 2002).

in contexts that either satisfied the presupposition or did not. In the latter case, reading was slower and the effect was confined to the region containing the presupposition trigger and the author concluded that presuppositions play an important role in online sentence comprehension. Tiemann et al. (2011) investigated five different triggers in a self-paced reading experiment. Participants were instructed to read a context and a test sentence in three different conditions, one including a presupposition trigger, one with a neutral word that triggered no presupposition, and one with a word that made the sentence ungrammatical. They also observed increased reading times already on the trigger and concluded that processing presuppositions starts already on the trigger. Recent studies (e.g., Domaneschi & Di Paola, 2018; Domaneschi et al., 2014; Tiemann et al., 2015) suggest that different types of presupposition triggers differ in processing, and thus it is unfortunate that Tiemann et al. (2011) did not report separate analyses. Schneider, Bade, and Janczyk (2020) focused on only two triggers (definite determiners and *again*) and basically replicated the results of Tiemann et al. (2011) in their Experiment 1 and employed separate analysis for the two different triggers. Further, their Experiment 2 employed a similar method as the present paper does to investigate whether presupposition verification is capacity-limited or not. (More details will be provided after introducing the employed method.)

Further evidence for additional processing difficulties if the determiner is used infelicitously against the provided context comes from a self-paced reading study by Altmann and Steedman (1988). These authors investigated how sentences with a definite determiner are processed if the uniqueness-presupposition is not met and observed that test sentences with an unmet uniqueness-presupposition were read slower. Van Berkum, Brown, Hagoort, and Zwitserlood. (1999; see also van Berkum, Brown, & Hagoort, 2003) also investigated referentially ambiguous noun phrases and observed early effects when the uniqueness presupposition was not met. Additional evidence from a study utilizing event related potentials (ERPs) by Burkhardt (2006). A test sentence with a definite noun phrase was presented in different contexts that either explicitly introduced the mentioned individual (*given* condition), made it easy to infer the mentioned individual (*bridged* condition), or made it impossible to infer the mentioned individual (*new* condition). The results also revealed a negative shift that emerged in the ERPs at about 300–400 ms after acoustic onset of the noun (see van Berkum et al., 2003) when the existence-presupposition of the definite determiner was not given.

Differences in processing of both determiners were also investigated. For example, Kirsten et al. (2014) used ERPs to investigate processing of definite and indefinite determiners in contexts that either fulfilled the uniqueness-assumption or did not. The nonmatching determiner evoked early effects (N400/P600) immediately on reading the determiners. The enhanced negativity in case of the definite compared with the indefinite determiner at least suggests that the underlying processes between the two determiners differ. Similar results were obtained in a MEG study with auditorily presented test sentences (Hertrich et al., 2015).

Some studies also reported more difficult processing of the indefinite compared to the definite determiner (Schneider, Schonard, Franke, Jäger, & Janczyk, 2019; Schumacher, 2009). Schneider et al. (2019) directly investigated the two-step evaluation that follows from the anti-presupposition account of the indefinite determiner. To this end, they used mouse-tracking to compare processing of both determiners in felicitous and infelicitous sentences, as well as in logically false sentences. In two experiments, participants were asked to judge the appropriateness of sentences like "Of these, Jan received the/a banana." in contexts showing that Jan's mother bought three pieces of fruit (e.g., one banana and two pears), and handed one of these fruits to Jan. Participants made their judgment via moving the mouse cursor into response boxes located in the top left and right corner of the computer screen. In the false conditions, the depicted fruit and the fruit mentioned in the test sentence did not match (e.g., the picture depicted Jan with a pear and the test sentence said "Jan received the/a banana"). For the felicitous and infelicitous conditions, they matched, but the determiner was used felicitously or infelicitously against the presented context. Mouse-tracking gives insight into how processing the definite and indefinite determiner unfolds over time (a similar procedure was also used by Tomlinson, Bailey, & Bott, 2013, to investigate the processing of scalar implicatures). In both felicitous and infelicitous sentences, the indefinite determiner was more difficult to process than the definite determiner. Most importantly, mouse-trajectories also revealed an initial deviation toward the response that was eventually not chosen (i.e., toward the non-target) in felicitous and infelicitous sentences for the indefinite determiner (i.e., larger "X_{neg} values"[2]). This was predicted by the two-step account: (1) participants initially consider the uniqueness-presupposition (2) before then negating this presupposition. Thus, these data can be taken as evidence for the anti-presupposition nature of the anti-uniqueness inference. At the same time, the data did not reveal a clear difference between the felicitous and infelicitous uses, even though the descriptive statistics of all dependent measures were in the predicted direction, that is, larger measures for infelicitous compared to felicitous uses (see also Clifton, 2003, Exp. 1 and 3). The authors suggested that this unexpected observation may have resulted from including the logically false sentences that potentially introduced some response ambiguity: First, the 'false' response was required for both the infelicitous and the false sentences, and second, a bias toward the 'false' response may have emerged, because this response was required in two-thirds of the trials.

In sum, the literature suggests that presuppositions that are supported by the present context are indeed processed immediately. Furthermore, additional processing costs in cases of infelicitously used presupposition triggers, and more complex processing of indefinite compared to definite determiners can be concluded from the available literature. In the present experiment we focus on the verification of presuppositions in case of determiners and ask whether the implicated processes are capacity-limited or can run in parallel with other ongoing processes.

[2] X_{neg} is the amount of horizontal deviation toward the competitor (i.e., toward the non-target response) and is measured as the maximum x-coordinate reached by the mouse as it veers away from the target response. This measure is particularly useful for detecting two-step processes, in which the second step involves a negation of the first step and is therefore expected to trigger a reversal in mouse direction (cf. Tomlinson et al., 2013, for an example with implicatures).

Assessing capacity limitations: The Psychological Refractory Period (PRP) approach. The term "capacity-limited" is here used in a broad sense meaning that a process is capacity-limited if it is not automatic. In other words, a process is capacity-limited if its efficiency suffers (e.g., its duration increases) when it must run simultaneously with other processes. This is often translated in a way that a capacity-limited process requires the "central stage" and thus cannot run in parallel with other processes (see below where we introduce particular models). An accepted way to determine capacity limitations in cognitive processing uses the PRP approach in combination with the locus of slack-logic. Because of the complexity of this method, we first introduce the logic of this approach in general and then illustrate it with an example by Mädebach, Kieseler, and Jescheniak (2017).

In general, a PRP experiment is a variant of a dual task in which each trial consists of two different tasks that temporally overlap. The two stimuli (S1 and S2) are presented consecutively with a varying stimulus onset asynchrony (SOA) and require separate responses (R1 and R2). Typically, SOA has little or no influence on the response time in Task 1 (RT1), but those in Task 2 (RT2) increase when SOA decreases: the PRP effect (Telford, 1931; see also Fischer & Hommel, 2012; Janczyk, Humphreys, & Sui, 2019; Magen & Cohen, 2010; Miller, Ulrich, & Rolke, 2009; for exceptions, please see Janczyk, Wallmeier, Pfister, & Kunde, 2014). This pattern is often accounted for by the central bottleneck model (CBM; Pashler, 1984, 1994; Welford, 1952). The CBM assumes that processing of a task is divided into three stages, namely (i) a pre-central (perceptual) stage, (ii) a central stage, and (iii) a post-central (motor) stage (see also Fig. 1a). While pre- and post-central stages can be processed in parallel with other stages, the critical assumption concerns the central stage, which is conceived capacity-limited and only one central stage can be processed at any time, thus constituting a bottleneck.[3]

[3] There are, of course, other models addressing dual-task costs and the PRP effect (for a recent overview, see Koch, Poljac, Müller, & Kiesel, 2018). For example, while the CBM describes the bottleneck as structural, other authors viewed the bottleneck as a strategic implementation to avoid crosstalk between tasks (e.g., Meyer & Kieras, 1997). In contrast, capacity-sharing models assume that central processes can run in parallel, but become less efficient then (Navon & Miller, 2002; Tombu & Jolicoeur, 2003). In addition, the nature of the capacity-limited process

As a consequence, with a short SOA, the central stage of Task 2 must be postponed until the central stage of Task 1 has finished. This waiting time is called *cognitive slack* and increases the corresponding RT2. In contrast, with a long SOA, the central stage of Task 1 is already processed and no (or only little) cognitive slack occurs, leading to shorter RT2 (see Fig. 1a).

Within a PRP experiment, the locus of slack-logic (Schweickert, 1978) allows to identify whether an RT effect (1) results from the parallel, pre-central stage of processing or (2) from the capacity-limited central stage of processing.[4] To this end, the manipulation yielding the RT effect is implemented in Task 2.

Consider first the case where a manipulation in Task 2 affects processing during the capacity-limited central stage of Task 2, the duration of which then becomes longer (see Fig. 1b). With a long SOA, the manipulation prolongs the central stage and thereby RT2 becomes longer. Similarly, with a short SOA, processing of the central stage of Task 2 has to wait until the central stage of Task 1 has finished. The effect caused by the manipulation in Task 2 prolongs the central stage of Task 2 and thus RT2 as well. In other words, the same RT2 difference is observed with a short and long SOA which statistically means an additive combination of SOA and the manipulation in Task 2 (i.e., no interaction).

Now consider the case where the manipulation affects parallel, pre-central processing (see Fig. 1c). With a long SOA, the manipulation becomes evident in an RT2 difference as well. With a short SOA, in contrast, the longer duration of pre-central processing runs while the central stage of Task 1 is still processed. Therefore, the effect is absorbed into the cognitive slack and there will not be an observable difference for RT2. Statistically, the manipulation and SOA interact in an underadditive way.

---

and its meaning for dual-task costs is a topic being discussed (Janczyk & Kunde, 2020). It is finally worth a mention that bottleneck and capacity sharing models make the same predictions for the critical Task 2 results.

[4] To be precise, it cannot be distinguished with this logic whether an effect results from the central or the post-central stage of processing. To do so, the effect propagation-logic might be used (e.g., Durst & Janczyk, 2018; Janczyk et al., 2019; Miller & Reynolds, 2003).

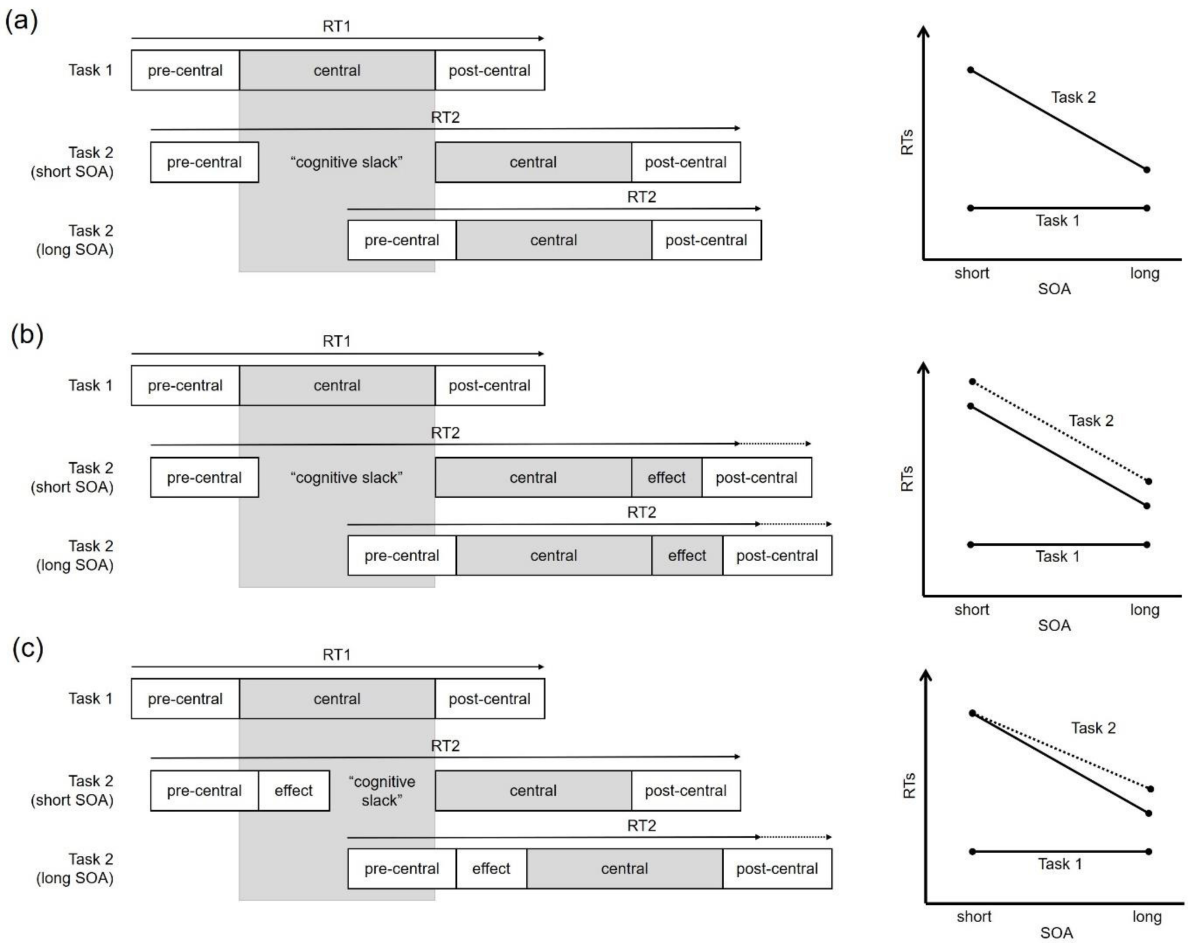


Figure 1. Illustration of the central bottleneck model (a) and the predictions of the locus of slack-logic (b) and (c). In (b) and (c), the solid and dotted lines for Task 2 represent two levels of the manipulation implemented in Task 2.

This method was, for example, used to investigate during which processing stage semantic interference effects arise in (Mädebach, Kieseler et al., 2017). In a previous study, Mädebach, Wöhner, Kieseler, and Jescheniak (2017) observed interference between semantically related environmental distractor sounds and picture naming. When, for example, naming a "horse", the effect was larger when the distractor sound came from a semantically related environment (e.g., a barking dog) compared to unrelated environmental distractor sounds (e.g., drums). Mädebach, Kieseler et al. (2017) then investigated the processing stage responsible for this effect and thus aimed to distinguish between the competitive selection view (e.g., Levelt, Roelofs, & Meyer, 1999) and the non-competitive selection view (e.g., Mahon, Costa, Peterson, Vargas, & Caramazza, 2007). According to the former, the effect would arise during response selection in picture naming and thus be a capacity-limited process. In contrast, the effect would arise during preselection processes like visual processing and object identification and thus be an automatic process according to the latter view.

In their experiment, Mädebach, Kieseler et al. (2017) had participants to perform a geometric form classification task with manual responses as Task 1, and a picture-sound interference task with vocal responses as Task 2. Stimuli in both tasks were presented with an SOA of 0 versus 500 ms. Critically, pictures in Task 2 were paired with different sounds resulting in different conditions. We will focus on two conditions that are relevant to explain the method. In the *semantic condition*, the picture was paired with a sound from the same semantic category (e.g., picture "horse" with a barking sound from a dog). In the *unrelated condition*, the picture was paired with an unrelated sound (e.g., picture "horse" with a drumming sound from a drum). Considering only the long SOA of 500 ms, the first prediction is that RT2 should be longer in the semantic compared with the unrelated condition.

For the short SOA, in contrast, both theoretical views make different predictions. First, if the picture-sound interference effect arises from the capacity-limited selection stage, a difference in RT2 of the same size as with the long SOA is expected (cf. Fig. 1b). In other words, SOA and picture-sound interference combine additively. Second, if the interference effect arises from preselection processes, the additional processing time becomes absorbed into the cognitive slack and no RT2 difference is observed (cf. Fig. 1c), and an underadditive interaction of SOA and picture-sound interference is predicted.

The data clearly revealed an additive combination of SOA and picture-sound interference, suggesting that the picture-sound interference effect requires central capacity and arises from the central stage. Such pattern has also been observed for, for instance, semantic interference in picture naming (Piai, Roelofs, & Schriefers, 2013), encoding into short-term memory (Jolicoeur & Dell'Acqua, 1998), switching between working memory items (Janczyk, 2017), and resolving Garner interference (Janczyk, Franz, & Kunde, 2010).

Capacity limitations and presupposition processing: The present study. We here apply the PRP approach and the locus of slack-logic to presupposition processing. Task 1 in our experiment was a tone discrimination task. Task 2 was a sentence evaluation task borrowed from Schneider et al. (2019). In this task, participants were provided with a context and asked to evaluate the appropriateness of a test sentence against this context (see Fig. 2 for an illustration; a more detailed description is provided in the Method section). This combination brings together well-established cognitive psychologist's chronometric tools with a topic from the field of pragmatic. We believe that such cross-domain research can stipulate methodological advance and also yield important results for linguistic theory building.

Apart from the methodological contribution, we pursued three goals with our experiment. The first goal was to replicate the more difficult processing of indefinite compared with definite determiners with RTs instead of mouse-trajectory analyses (as in Schneider et al., 2019).

The second goal was to investigate whether processes implicated in interpretations of the indefinite determiners (1) run in parallel with other central stages or (2) are capacity-limited. Considering the indefinite determiner as an anti-presupposition suggests the longer RTs for the indefinite compared to the definite determiner to result from negating the uniqueness-presupposition of the definite determiner. Processing linguistic negations was shown to induce longer RTs than processing affirmative sentences due to two-step process where the to-be-negated information is activated first and subsequently negated (Dudschig & Kaup, 2018). Further, negation was argued to be a "resource dependent process" (Strack & Deutsch, 2004, p. 227; see, e.g., Deutsch, Kordts-Freudinger, Gawronski, & Strack, 2009, Exp. 3; Foerster, Wirth, Berghoefer, Kunde, & Pfister, 2019; Wason, 1959). Thus, inferring anti-uniqueness should require central capacity and thus an additive combination with SOA is predicted. In other words, the RT difference between definite and indefinite determiners is expected to be the same at both SOAs with no interaction of determiner and SOA. This expectation receives further preliminary support from a recent study with a self-paced reading experiment (Schneider et al., 2020). In this study, processing costs of the definite determiner and the presupposition trigger *again* when compared to non-trigger words were investigated. The results of this study are more in line with the idea that presupposition processing requires limited capacities, but it is important to note that this is a different question than we ask here for the comparison between the definite and the indefinite determiner.

The third goal was to re-address the unexpected observation of Schneider et al. (2019) that infelicitous sentences were not more difficult to process than felicitous ones were. Schneider et al. reasoned that this observation may have resulted from peculiarities of the particular design. We here eliminated these problematic aspects by omitting the "logically false" sentences. If the reasoning were true, we would then expect longer RT2s for the infelicitous sentences (see, e.g., Burkhardt, 2006; Kirsten et al., 2014; Tiemann et al., 2011; see also Clifton, 2013, for the possible role of a dual-task). In addition, we expected this RT difference to also combine additively with SOA, indicating the implication of limited central capacities. In other words, no interaction of sentence type and SOA is predicted and the differences between felicitous and infelicitous sentences should be the same for both SOAs. More precisely, the presupposition trigger is assumed to immediately start a context search for an appropriate referent introduced by the preceding context. This search process in turn involves repeated selection and de-selection of working memory items that are considered as referents. Janczyk (2017) observed that this selection and de-selection prosses is a capacity limited. In case of an infelicitous sentence, detecting the mismatch requires more (and unsuccessful) comparisons, but no appropriate referent is found and finally the search process has to be terminated.

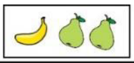
Method

Participants. An a-priori power analysis for a paired *t*-test (δ=0.5, 1-β=.9, α=.05) yielded a minimum sample size of $n = 44$, and thus our intended sample size was $n = 48$. To achieve this, data were originally collected from 54 native speakers of German from the Tübingen (Germany) area, of which six participants were excluded because of more than 30% errors (final sample: mean age = 22.1 years, 38 females, 10 males). All participants were native speakers of the German language, reported normal or corrected-to-normal vision and hearing abilities, and signed written informed consent prior to data collection.

Apparatus and Stimuli. Stimulus presentation and response collection were controlled by a standard PC connected to a 17-in. CRT monitor. Two external response keys each to the left and to the right of the participants were used for response collection, and were operated with the index- and middle-fingers of both hands. Responses with the left hand were required in Task 1, and with the right hand in Task 2. Auditory stimuli for Task 1 (S1) were 300 and 900 Hz sinusoidal tones (50 ms) presented via headphones. Visual stimuli for the sentence evaluation task (Task 2) were adopted from Schneider et al. (2019; for an example see Fig. 2). A rectangle with three items of fruit was used as the shopping basket of "Jan's mother", and a picture of "Jan" was visible below. Sentential stimuli (i.e., the context and the test sentence) were presented in black font on the otherwise white background. Seven kinds of fruit (banana, lemon, orange, pear, pineapple plum, and strawberry) were used in the experiment. On each trial, the shopping basket comprised two items of one fruit and one different item that were introduced as the current context. All test sentences were of the form "Of these, Jan received [determiner: *the* or *a*] [fruit].", and only the feminine determiner was used in conjunction with the seven kinds of fruit to avoid the possibility of early disambiguation. To facilitate interpretation of the test sentence and evaluation of the determiner according to the current context, test sentences started with "Davon…" (English *Of these, …*). This emphasizes that Jan received the fruit from the three initial fruit of the shopping basket and minimizes the possibility of interpreting the determiner according to a global context (e.g., all available fruit in the world).

This procedure is thought to maximize the likelihood of interpreting the determiner in the restricted local context (Singh, 2011).

Four sentence types resulted from combining the definite and the indefinite determiner with their felicitous and infelicitous uses (see Fig. 2 for an overview). With seven types of fruit it is possible to create 42 felicitous and 42 infelicitous combinations with definite and indefinite determiners, resulting in 168 possibilities.

Context: Jan's mother was shopping. She bought one banana and two pears. 

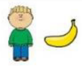
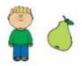
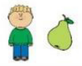
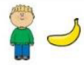
Sentence type	Determiner	Picture	Stimulus sentence	Expected response
felicitous	definite		Of these, Jan received **the** banana.	green
	indefinite		Of these, Jan received **a** pear.	green
infelicitous	definite		Of these, Jan received **the** pear.	red
	indefinite		Of these, Jan received **a** banana.	red

Figure 2. Examples for the four different conditions resulting from combining the two determiners (definite vs. indefinite) with two sentence types (felicitous vs. infelicitous).

We did not include filler items to avoid a reduction in the amount of test items. Including filler items bears the risk of more noisy data and worse point estimation. In this, we also followed published work on scalar implicatures (Bott & Noveck, 2004) and processing of negation (Dale & Duran, 2011), and additionally we aimed to keep these procedural aspects close to the study of Schneider et al. (2019).

Task and Procedure. Task 1 was to respond to the pitch of a tone (S1) with a left-hand key press (R1). Task 2 was to indicate the appropriateness of the test sentence (S2) against the current context with a right-hand key press (R2) (e.g., “green” in the example in Fig. 3).

Each trial started with a picture of Jan and a context in the form of the shopping basket with three pieces of fruit (the context relevant fruit) and an additional sentential description (for example, ”Jan's mother was shopping. She bought one banana and two pears.”; Fig. 3.1). After

1000 ms, Jan received one piece of the context fruit, additionally visualized by an arrow to emphasize the transfer to Jan. Further, the first part of the test sentence became visible as well ("Of these, Jan received…"; Fig. 3.2). Pressing both left-hand response keys simultaneously initiated the next step, and S1 was played after 500-700 ms (randomly determined; Fig. 3.3). Following an SOA of 100 or 1200 ms, the test sentence was completed on the screen as S2 (e.g., "Of these, Jan received a pear."; Fig. 3.4).

In case of errors in either task, written feedback ("Wrong answer in tone discrimination task." or "Wrong answer in sentence evaluation task.") was provided on the screen for 2000 ms. Further error feedback was provided when participants did not perform according to the instructions (unspecific errors; i.e., when they performed Task 2 before Task 1, responded twice in one task, or did not respond within 5000ms). The next trial started after an inter-trial interval (ITI) of 1000 ms. Task 2-related feedback was also applied to bias participants towards the intended interpretation of the German word "eine" as an indefinite determiner (see Schneider et al., 2019, for a justification, and also Bott & Noveck, 2004, or Tomlinson et al., 2013, for a similar approach with scalar implicatures).

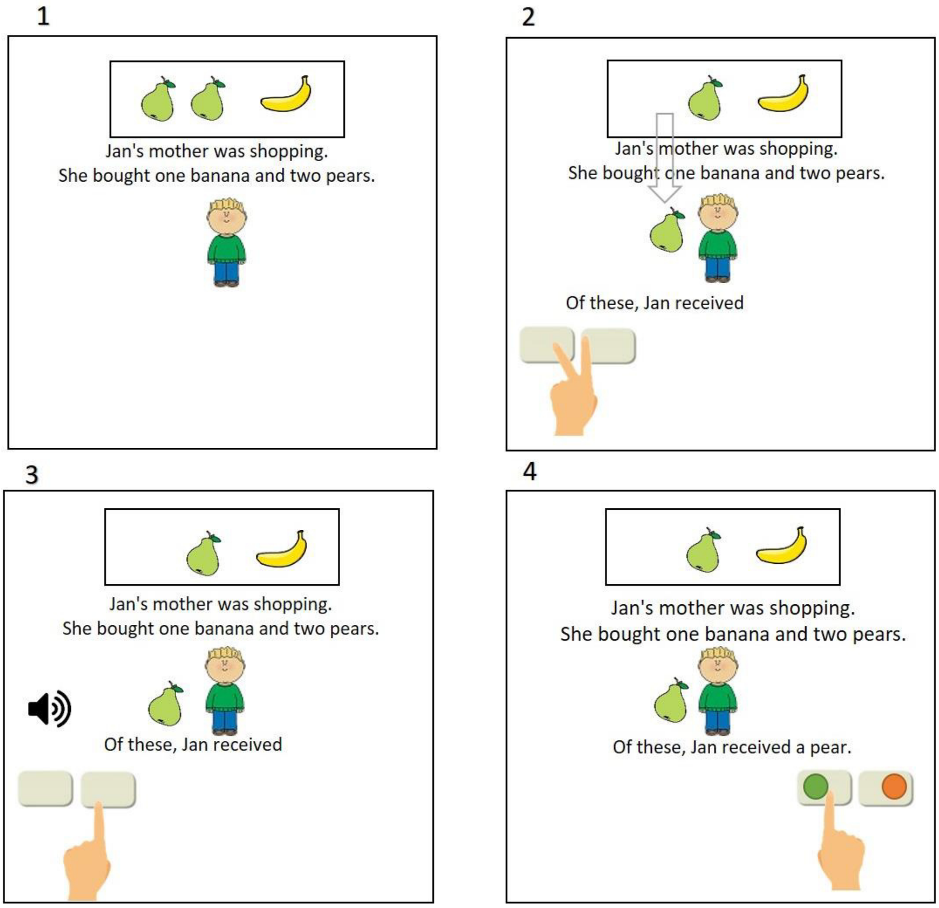


Figure 3. Illustration of the trial procedure (see text for more information).

Participants were tested individually in a single session of about 30 minutes. A session started with 40 practice trials of only Task 1. These trials were initiated by written instructions centered on the screen asking participants to press both left-hand response keys simultaneously to start a trial (to mimic the procedure required in the main experiment). After 100 ms, S1 was played (each pitch 20 times in a random order) and participants were to respond with a left-hand key press. In case of errors, respective error feedback was presented on the screen (1000 ms). The ITI was 1000 ms. After practicing Task 1, two unanalyzed practice blocks including both tasks with six trials of each condition (i.e., 48 practice trials in total; resulting from using the same six trials in four conditions in both practice blocks) were administered. Following that, three experimental blocks with twelve trials per condition (i.e., 48 trials per block, or 144 trials in total) were administered. In sum then, all logically possible combinations were used. Each participant saw six practice items (twice in each practice block) and 36 experimental items in each of the four conditions, resulting in the 168 available items. The trials were administered in random order.

Participants received written instructions before the experiment that were also visually illustrated with an example trial. Instructions focused on speed while maintaining a low level of errors. Stimulus-response mappings were counterbalanced across participants.

Design and Analysis. Dependent measures were RTs and percentage of errors (PEs) in Task 1 and Task 2. For all analyses, trials with unspecific errors were excluded first (3.24%). The most important RT analyses were based on correct trials only and trials with RTs deviating more than 2.5 *SD*s from the mean of the respective cell (calculated separately for each participant) were excluded as outliers. To exclude speed-accuracy tradeoffs (see Liesefeld & Janczyk, 2019), we also analyzed PEs in the same way as RTs (what is standard in RT research). In particular, RTs and PEs in Task 2 and Task 1 were submitted to a 2 × 2 × 2 ANOVA with SOA (100 vs. 1200 ms), determiner (definite vs. indefinite), and sentence type (felicitous vs. infelicitous) as repeated-measures. Because the critical results concern the additivity of determiner and sentence type with SOA (i.e., non-significant two-way interactions), these interactions were also assessed with Bayes factors calculated with the R-package BayesFactor (Morey & Rouder, 2018). Bayes factors BF_{01} were calculated as the ratio of (1) the model including the three main effects and the interaction of interest and of (2) the model including only the three main effects, and they are reported following the respective *F*-test of the ANOVA. In addition, although such ANOVA-approach has been used in previous psycholinguistic PRP experiments as well (Mädebach, Kieseler et al., 2017; Piai et al., 2014), Task 2 RTs and PEs were additionally analyzed with a (generalized) linear mixed effect model with participant and stimulus/context-set (i.e., item) as random effects, and the results are reported in the Appendix. All analyses were carried out using the R software (R Core Team, 2020).

Results

Because the theoretically most important results relate to Task 2, we begin with these results and the Task 1 results are reported afterwards.

Task 2. Mean RT2s (2.76% outliers) are summarized in Table 1. RTs were longer with a short than with a long SOA (1502 vs. 784 ms), thus a PRP effect, $F(1,47) = 1183.87$, $p < .001$, $\eta_p^2 = .96$. Further, RT2 was longer for indefinite compared to definite determiners (1163 vs. 1123 ms), $F(1,47) = 10.34$, $p = .002$, $\eta_p^2 = .18$, and in infelicitous compared to felicitous sentences (1098 vs. 1188 ms), $F(1,47) = 53.41$, $p < .001$, $\eta_p^2 = .53$ (see Fig. 4 for an illustration of these main effects). Most importantly, determiner did not interact with SOA, $F(1,47) < 0.01$, $p = .948$, $\eta_p^2 < .01$, $BF_{01} = 3.17$, nor did sentence type interact with SOA, $F(1,47) = 1.23$, $p = .273$, $\eta_p^2 =.03$, $BF_{01} = 6.49$. However, determiner and sentence type interacted, $F(1,47) = 15.92$, $p < .001$, $\eta_p^2 = .25$: For felicitous sentences, RT2s were longer for indefinite compared with definite determiners (1143 vs. 1053 ms; $\Delta = 90$ ms), but the difference was reversed and small in size for infelicitous sentences (1182 vs. 1193 ms; $\Delta = -11$ ms; see Fig. 5 for an illustration). The three-way interaction was not significant, $F(1,47) = 0.80$, $p = .374$, $\eta_p^2 =.02$.

Mean PEs are summarized in Table 1. Only the main effect of determiner was significant, $F(1,47) = 19.85$, $p < .001$, $\eta_p^2 =.30$, with more errors for indefinite compared to definite determiners. All other Fs $\leq$ 2.77, all ps $\geq .103$.

Table 1. Mean RTs (in milliseconds, ms) | PEs (percentage of errors) for the sentence evaluation task (Task 2) as a function of determiner, sentence type, and stimulus onset asynchrony (SOA).

	determiner			
	definite		indefinite	
	SOA [ms]		**SOA [ms]**	
sentence type	100	1200	100	1200
felicitous	1401 | 4.71	705 | 3.87	1500 | 8.12	786 | 8.40
infelicitous	1563 | 6.20	824 | 5.96	1544 | 9.28	819 | 8.82

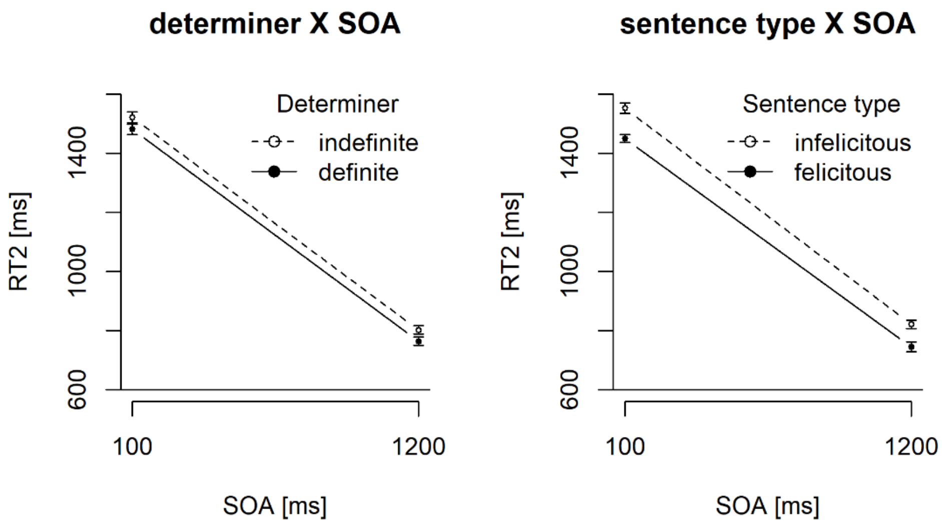


Figure 4. Response times for the sentence evaluation task (RT 2) as a function of determiner and stimulus onset asynchrony (SOA) (left) and (b) sentence type and SOA (right). Error bars are within-subjects standard errors (Morey, 2008).

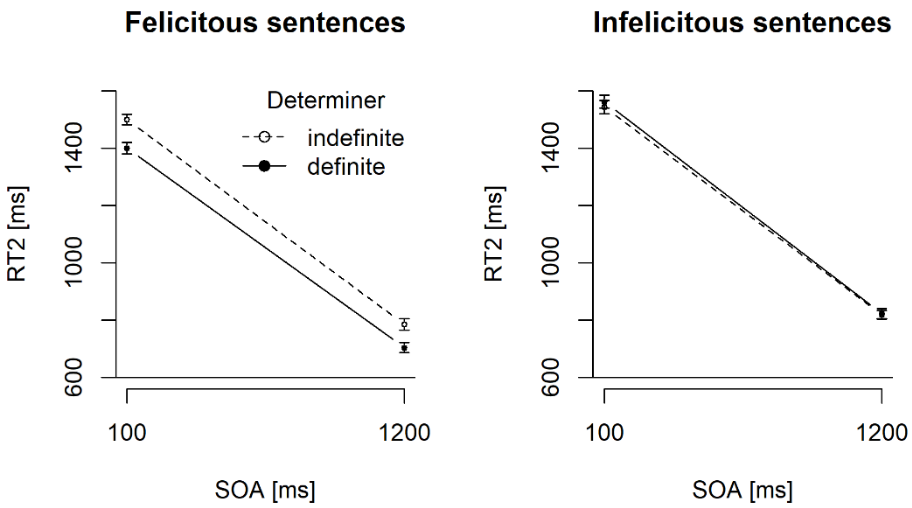


Figure 5. Response times for the sentence evaluation task (RT 2) as a function of determiner and stimulus onset asynchrony (SOA) separately for felicitous (left) and infelicitous sentences (right). Error bars are within-subject standard errors (Morey, 2008).

Task 1. Mean RT1s (2.82% outliers) and mean PE1s are summarized in Table 2. RT1s were longer with a short compared to a long SOA (928 vs. 773 ms), $F(1,47) = 73.91$, $p < .001$, $\eta_p^2 = .61$, and were longer for infelicitous compared to felicitous sentences (861 vs. 840 ms), $F(1,47) = 4.78$, $p = .034$, $\eta_p^2 = .09$. The main effect of determiner was not significant (849 vs. 852 ms, definite vs. indefinite determiner, respectively), $F(1,47) = 0.17$, $p = .686$, $\eta_p^2 < .01$. The effect of sentence type was larger at the short SOA (952 vs. 904 ms; $\Delta = 48$ ms) than at the long SOA (771 vs. 776 ms; $\Delta = -5$ ms), thus an overadditive interaction, $F(1,47) = 7.88$, $p = .007$, $\eta_p^2 = .14$. Determiner did not interact with SOA, $F(1,47) = 0.30$, $p = .585$, $\eta_p^2 = .01$, nor with sentence type, $F(1,47) = 3.46$, $p = .069$, $\eta_p^2 = .07$. The three-way interaction was significant, $F(1,47) = 5.38$, $p = .025$, $\eta_p^2 = .10$. One way to view this interaction is that the overadditive interaction of sentence type and SOA was more clearly expressed for definite than for indefinite determiners. For PE1s, no effect reached significance, all Fs ≤ 2.46, all ps $\geq .123$.

Table 2. Mean RTs (in milliseconds, ms) | PEs (percentage of errors) for the auditory discrimination task (Task 1) as a function of determiner, sentence type, and stimulus onset asynchrony (SOA).

	determiner			
	definite		indefinite	
	SOA [ms]		**SOA [ms]**	
sentence type	100	1200	100	1200
felicitous	888 |1.17	780 | 0.46	931 | 0.72	769 | 0.35
infelicitous	958 | 0.72	773 | 0.93	950 | 1.31	769 | 0.70

Discussion

The present study investigated whether processing the (anti-) presuppositions of determiners requires limited cognitive capacities or proceeds rather automatically. To this end, we embedded a sentence evaluation task (see Schneider et al., 2019) as Task 2 in a PRP experiment with an auditory discrimination task as Task 1 (see Pashler, 1994, for a review).

Summary of results. The two main effects that were in the focus of the present study support our hypotheses. First, processing sentences with indefinite determiners took longer compared to those with definite determiners. This result conceptually replicates the observations by Schneider et al. (2019; see also Hertrich et al., 2015; Kirsten et al., 2014; Schumacher, 2009) and suggests that indeed processing indefinite determiners is more difficult and/or requires an extra processing step compared with the definite determiner. Second, participants also responded more slowly when the determiner was used infelicitously against the current context compared to its felicitous use. Such result has been reported in previous studies (Altmann & Steedman, 1988; Clifton, 2003, Exp. 2 and 4; van Berkum et al., 1999, 2003), but failed to replicate in Schneider et al. (2019). Both effects were of the same size with the long and with the short SOA, that is, they combined additively with the SOA manipulation. Although this conclusion is based on retaining the null-hypotheses of interactions, additional Bayesian analyses revealed $BF_{01} > 3.0$ in both cases, thus strengthening the evidence for the null-hypotheses.

However, the two main effects need an important qualification, because determiner and sentence type interacted. Essentially, the longer RTs with indefinite compared with definite determiners was only visible in felicitous sentences, while RTs were roughly the same in infelicitous sentences. At the same time, error rates were higher for indefinite compared with definite determiners, irrespective of sentence type.

Theoretical implications and limitations. The observed RT difference between the definite and the indefinite determiner in the felicitous sentences, fits well with the anti-presupposition account for the anti-uniqueness association of the indefinite determiner, as derived from application of the Maximize Presupposition principle (Heim, 1991, 2011): Evaluation of the indefinite determiner involves first evaluating the uniqueness-presupposition of the definite determiner and in a second step its negation. Thus, the indefinite determiner does not presuppose anti-uniqueness itself. We concur that the assumed negation was not directly manipulated in our experiment, but still it was assumed as the reason for the increased processing difficulties for the indefinite determiner and the results are in agreement with this assumption. However, the two-step nature of interpreting the indefinite determiner becomes visible more directly in the mouse-trajectory analyses of Schneider et al. (2019) where especially high X_{neg} scores revealed that participants initially considered the competitor response for indefinite determiners. The additivity with SOA means, in terms of the locus of slack-logic, that one or both processes require the central bottleneck and thus require limited cognitive capacities. In fact, the assumed second process of negation has been suggested as a capacity-limited process (see Strack & Deutsch, 2004). Thus, this aspect of the results first replicates previous results and thereby strengthens the belief in more difficult processing of the indefinite compared with the definite determiner. In addition, the conclusion that the additional process responsible for the processing difficulties is capacity-limited helps delineating the nature of this process: negation of the uniqueness-presupposition of the definite determiner indeed appears to be a viable candidate.

Further, these results would also be in line with the assumption that presupposition triggers start a context search for an appropriate referent. This search process likely involves working memory resources. An involvement of working memory in presupposition processing has been suggested from different perspectives as well. This makes sense against the present results, because working memory has often been described as severely limited in capacity. That such kind of cognitive work load influences presupposition processing was reported by Domaneschi et al. (2014). In this study, participants had to answer yes/no questions on presuppositional content of a presented context and cognitive work load was manipulated with a memory task. The study revealed that the cognitive demand of presuppositions is related to the complexity of the mental representation that the presupposition requires. Definite descriptions require a simple representation of a referent in the mental model, whereas change of state verbs seem to be more complex and require different temporal representations. Further evidence for an involvement of working memory in presupposition processing can be derived from a study by Anderson and Holcomb (2005), who observed an increased left anterior negativity (LAN) for definite compared with indefinite determiners, what can be interpreted as a referential assignment increasing demands on working memory (see also King & Kutas, 1995). Additional support for an involvement of working memory resources in presupposition processing comes from a recent aging study. Di Paola and Domaneschi (2019) investigated how the decline of working memory abilities with increasing age affects presupposition processing and observed longer RTs for older participants, especially for presuppositions triggered by change of state verbs like *stop*. In their conclusion, the authors argued that change of state verbs require a more demanding mental representation (as, e.g., definite descriptions) and thus more cognitive recourses are used. The decline of working memory abilities with age has an effect on processing of presuppositions and can explain why older participants need more time.

Why did we not observe a similar RT effect in infelicitous sentences? Tentatively we can offer two explanations for this unexpected observation. First, to some degree a speed-accuracy tradeoff (e.g., Liesefeld & Janczyk, 2019) might have contributed, because the error rate was consistently larger for the indefinite determiner across all conditions. Second, Bade and Schwarz (2019) observed that the anti-uniqueness inference of the indefinite determiner is evaluated less often than the uniqueness-presupposition of the definite determiner. This can explain why participants make more errors in the infelicitous indefinite condition, but it appears difficult to explain RT patterns based on this argument. Third, a possible explanation for the absence of an RT effect between the two determiners in the infelicitous condition might be that the definite determiner is even more implausible in a situation depicting multiple objects than the indefinite determiner is in the context of a single object. This might in fact lead to a larger effect of infelicity for the definite compared to the indefinite determiner, increase the corresponding RTs, and mask the otherwise expected RT difference between the two determiners.

It certainly deserves a mention that this aspect of the results was not expected: Schneider et al. (2019) obtained the disadvantage for indefinite compared with definite determiners for both felicitous and infelicitous sentences, though not in RTs proper but in measures derived from mouse trajectories. Whether peculiarities of the design or the different dependent measures played an important role for these diverging results is an interesting topic for future research. One possible candidate for a critical difference is the response mode. With key press responses—as used in the present study—one can only measure RTs at the final decision for which no correction is further possible. In contrast, mouse movements allow to observe unfolding of the decision-making process. With this method it is thus possible that participants change their mind before reaching the final response box. Taken together, longer movement times etc. in the Schneider et al. (2019) study may be a by-product of the initial deviation toward the wrong response in case of indefinite determiners, which is the main indicator for the assumed two-step processing though.

The longer RTs in infelicitous than in felicitous sentences clarifies the somewhat ambiguous results obtained by Schneider et al. (2019), although this effect was larger for the definite than the indefinite determiner. In the Schneider et al. study, this effect was not consistently significant, although descriptively in the expected direction, and the authors reasoned that the inclusion of logically wrong sentences might have worked against observing the expected effect. We here omitted these sentences and observed more difficult processing of infelicitous sentences. The additional processes running when evaluating an infelicitously used determiner also seem to be capacity limited. We suggest that encountering a presupposition trigger immediately starts a context search for an appropriate referent which likely involves repeated selection and de-selection of working memory items that are considered as referents. Such selection of working memory items has been shown capacity-limited (Janczyk, 2017).

In sum, the results support the notion that the processes initiated by presupposition triggers are capacity-limited rather than automatic. This conclusion receives additional evidence from a study by Schneider et al. (2020). This study combined the PRP approach with a self-paced reading task as used by Tiemann et al. (2011) and compared sentences including a presupposition trigger (definite determiner or *again*) with grammatical sentences without a presupposition trigger and unacceptable sentences. As expected, an RT difference between the trigger and the unacceptable condition was observed for both triggers. In combination with the SOA manipulation and the locus of slack-logic, the results are not in line with the idea that the trigger-initiated processing is automatic. More precisely, it was assumed that the search for an appropriate referent (in case of the definite determiner) or for a suitable previous event (in case of *again*) that is immediately started when encountering a presupposition trigger initially fails, because contrary to the current experiment, the context did not explicitly verify or falsify the presupposition. To prevent the sentence from being uninterpretable, participants anticipated accommodation (which was suggested by high acceptability ratings). Accommodation is understood as a process of enriching the context with the information of the presupposition (as long as this is contextually feasible; for a formal definition see Lewis, 1979). Before adding the presuppositional content to the context one has to check whether this is plausible, and this preparation for accommodation requires cognitive resources like working memory. It is important to note, however, that Schneider et al. (2020) investigated the processing of presupposition triggers in comparison to other words triggering no presupposition and words that made the sentence ungrammatical. It did not compare different triggers, nor did it investigate infelicitously used presupposition triggers. In contrast, the current experiment focuses on the comparison between the definite and the indefinite determiner and their felicitous or infelicitous use.

For Task 1 we observed a significant three-way interaction and an overadditive interaction of SOA and sentence type with longer RTs in long compared to short SOAs as well as in infelicitious compared to felicitous sentences. Against the background of the central bottleneck model (CBM), effects on Task 1 RTs are somewhat unexpected. However, such observations are not uncommon in PRP research. In particular, longer RT1 with decreasing SOA can easily be accounted for by central capacity sharing models (Tombu & Jolicoeur, 2003). Thus, while this part of the results can be taken to argue in favor of capacity sharing models of dual-tasking (see also Mittelstädt & Miller, 2017), these models make in fact the same predictions as the CBM regarding the critical Task 2 RTs (Navon & Miller, 2002). In addition, the effect and the overadditive interaction involving sentence type may be attributed to some trials with response grouping (Ulrich & Miller, 2008), that is, trials in which participants delayed responding in Task 1 until they have selected the Task 2 response as well.

The probably clearest limitation of the present study is its focus on the two determiners in the German language. These presupposition triggers were chosen, because the Maximize Presupposition principle was most often applied to determiners. Whether it can fruitfully be applied to other triggers as well is a case for future research. One candidate is “both” which has a competitor, namely “all”. In this case, Maximize Presupposition accounts for the fact that the possibility of using a sentence with “both” makes the sentence infelicitous when “all” is used. Thus, future studies should (1) investigate possible limits of the anti-presupposition approach and its implicated negation with other presupposition triggers, and (2) be carried out in different languages than German, to explore the generality of our results.

Conclusion. To conclude, the present experiment provides evidence for a more difficult processing of indefinite determiners, as would be expected if the anti-uniqueness inference arises from a negation of the uniqueness-presupposition of the definite determiner. Thus, this result supports the idea of the Maximize Presupposition principle. Further, using determiners infelicitously seems to induce processing costs. Finally, the most important and novel contribution of the present study is that presupposition processing requires limited cognitive capacities. It also demonstrates that bringing together cognitive psychologist’s (chronometric) tools with the emerging field of experimental pragmatic (Noveck, 2019) can be a fruitful and promising endeavor for future research.

References

Alonso-Ovalle, L., Menéndez-Benito, P., & Schwarz, F. (2011). Maximize presupposition and two types of definite competitors. *Proceedings of North East Linguistic Society Conference (NELS)*, *39,* 29-40.

Altmann, G., & Steedman M. (1988). Interaction with context during sentence processing. *Cognition, 30,* 191-238.

Anderson, J. E., & Holcomb, P. J. (2005). An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain and Language, 94,* 200-216.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences *Journal of Memory and Language, 51*, 437–457.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language, 98*, 159-168.

Chemla, E. (2009). Presuppositions of quantified sentences: Experimental data. *Natural Language Semantics, 17,* 299-340.

Clifton, C. J. (2013). Situational context affects definiteness preferences: Accommodation of presuppositions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 487-501.

Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science, 35*, 983–996.

Deutsch, R., Kordts-Freudinger, R., Gawronski, B., & Strack, F. (2009). Fast and fragile. A new look at the automaticity of negation processing. *Experimental Psychology, 56,* 434-446.

Di Paola, S., & Domaneschi, F. (2019). The aging factor in presupposition processing. *Journal of Pragmatics, 140,* 70-87.

Domaneschi, F., & Di Paola, S. (2018). The processing costs of presupposition accommodation. *Journal of Psycholinguistic Research*, 47, 483-503.

Domaneschi, F., Carrea, E., Penco, C., & Greco, A. (2014). The cognitive load of presupposition triggers: Mandatory and optional repairs in presupposition failure. *Language, Cognition and Neuroscience*, *29*, 136-146.

Dudschig, C., & Kaup, B. (2018). How does "not left" become "right"? Electrophysiological evidence for a dynamic conflict-bound negation processing account. *Journal of Experimental Psychology: Human Perception and Performance, 44,* 716-728.

Durst, M., & Janczyk, M. (2018). The motor locus of the no-go based backward crosstalk. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 1931-1946.

Fischer, R., & Hommel, B. (2012). Deep thinking increases task-set shielding and reduces shifting flexibilityin dual-task performance, *Cognition, 123*, 303-307.

Foerster, A., Wirth, R., Berghoefer, F. L., Kunde, W., & Pfister, R. (2019). Capacity limitations of dishonesty. *Journal of Experimental Psychology: General, 148,* 943-961.

Frege, G. (1892). On sense and reference. In P. Geach & M. Black (eds.), *Translations from the philosophical writings of Gottlob Frege (pp. 56-78).* Oxford: Blackwell.

Grice, P. (1975), Logic and conversation. In P. Cole & J. L. Morgan (eds.), *Speech acts*. (pp. 41–58). New York, NY: Academic Press,

Heim, I. (1991). Artikel und Definitheit. In A. v. Stechow, & D. Wunderlich *Semantik / Semantics. Ein internationales Handbuch der zeitgenössischen Forschung. An international handbook of contemporary research* (pp. 487-535). Berlin: De Gruyter Mouton.

Heim, I. (2011). Definiteness and indefiniteness. In C. Maienborn, K. von Heusinger, & P. Portner (eds.), *Semantics: An international handbook of natural language meaning* (pp. 996-1025). Berlin: De Gruyter Mouton.

Hertrich, I., Kirsten, M., Tiemann, S., Beck, S., Wühle, A., Ackermann, H., & Rolke, B. (2015). Context dependent impact of presuppositions on early magnetic brain responses during speech perception. *Brain and Language*, *149,* 1-12.

Janczyk, M. (2017). A common capacity limitation for response and item selection in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1690-1698.

Janczyk, M., & Kunde, W. (2020). Dual tasking from a goal perspective. *Psychological Review.* http://dx.doi.org/10.1037/rev0000222

Janczyk, M., Franz, V. H. & Kunde, W. (2010). Grasping for parsimony: Do some motor actions escape dorsal processing? *Neuropsychologia, 48,* 3405-3415.

Janczyk, M., Humphreys, G. W., & Sui, J. (2019). The central locus of self-prioritization. *The Quarterley Journal of Experimental Psychology, 72*, 1068-1083.

Jolicoeur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36,* 138-202.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7,* 376–395.

Kirsten, M., Tiemann, S., C. Seibold, V., Hertrich, I., Beck, S., & Rolke, B. (2014). When the polar bear encounters many polar bears: Event-related potential context effects evoked by uniqueness failure. *Language, Cognition and Neuroscience*, *29*, 1147-1162.

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin, 144,* 557-586.

Levelt,W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–75.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic, 8,* 339-359.

Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy tradeoffs (?). *Behavior Research Methods, 51*, 40–60.

Mädebach, A., Kieseler, M.-L., & Jescheniak, J.D. (20182017). Localizing semantic interference from distractor sounds in picture naming: A dual-task study. *Psychonomic Bulletin & Review, 25,* 1909-1916.

Mädebach, A., Wöhner, S., Kieseler, M.-L., & Jescheniak, J. D. (2017). Neighing, barking, and drumming horses—Object related sounds help and hinder picture naming. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 1629–1646.

Magen, H., & Cohen, A. (2010). Modularity beyond perception: Evidence from the PRP paradigm, *Journal of Experimental Psychology Human Perception & Performance, 36*, 395-414.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture–word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 503–535.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms. *Psychological Review*, *104*, 3-65.

Miller, J. Ulrich, R., & Rolke, B. (2009) On the optimality of serial and parallel processing in the psychological refractory period paradigm: Effects of the distribution of stimulus onset asynchronies, *Cognitive Psychology, 58,* 273-310.

Miller, J., & Reynolds, A. (2003). The locus of redundant-targets and nontargets effects: Evidence from the psychological refractory period paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 1126–1142.

Mittelstädt, V., & Miller, J. (2017). Separating limits on preparation versus online processing in multitasking paradigms: Evidence for resource models. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 89-102.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4,* 61-64.

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes factors for common designs.* https://cran.r-project.org/web/packages/BayesFactor

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology*, *44*, 193-251.

Noveck, I. (2018). *Experimental pragmatics: The making of a cognitive science*. Cambridge: Cambridge University Press.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 358–377.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220–244.

Percus, O. (2006). Antipresuppositions. In A. Ueyama (ed.), *Theoretical and empirical studies of reference and anaphora: Toward the establishment of generative grammar as an empirical science* (pp. 52-73). Japan Society for the Promotion of Science.

Piai, V., Roelofs, A., & Schriefers, H. (2014). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 147-165.

R Core Team (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing.* Vienna, Austria. URL https://www.R-project.org/.

Sauerland, U. (2008) Implicated presuppositions. In A. Steube (ed.), *Sentence and context. Language, context & cognition* (pp. 581-600). Berlin: Mouton de Gruyter.

Schlenker, P. (2012). Maximize presupposition and gricean reasoning. *Natural Language Semantics, 20*, 391-429

Schneider, C., Bade, N., & Janczyk, M. (2020). Is immediate processing of presupposition triggers automatic or capacity-limited? A combination of the PRP approach with a self-paced reading task. *Journal of Psycholinguistic Research, 49,* 247-273.

Schneider, C., Schonard, C., Franke, M., Jäger, G., & Janczyk, M. (2019). Pragmatic processing: An investigation of the (anti-)presuppositions of determiners using mouse-tracking. *Cognition, 193,* 104024.

Schumacher, P. B. (2009). Definiteness marking shows late effects during discourse processing: Evidence from ERPs. *Lecture Notes in Computer Science, 5847,* 91-106.

Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a Stroop task. *Journal of Mathematical Psychology, 18,* 105–139.

Singh, R. (2011). Maximize Presupposition! and local contexts. *Natural Language Semantics, 19*, 149–168.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy, 25*, 701–721.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8,* 220-247.

Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology, 14,* 1-36.

Tiemann, S., Kirsten, M., Beck, S., Hertrich, I., & Rolke, B. (2015). Presupposition processing and accommodation: An experiment on wieder ('again') and consequences for other triggers, in Schwarz (Ed.) *Experimental Perspectives on Presuppositions, vol 45* (pp. 39-65). Cham: Springer.

Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., Knapp, J., & Beck, S. (2011). Psycholinguistic evidence for presuppositions: On-line and off-line data. *Proceedings of Sinn & Bedeutung, 15*, 581-595.

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 3-18.

Tomlinson J. M., Bailey, T., Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, *69*, 18-35.

Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language, 69,* 18-35.

Ulrich, R., & Miller, J. (2008). Response grouping in the psychological refractory period (PRP) paradigm: Models and contamination effects. *Cognitive Psychology, 57,* 75-121.

van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology 40*, 235-248.

van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41,* 147-182.

Welford, A. T. (1952). The ‘psychological refractory period’ and the timing of high-speed performance—a review and a theory. *British Journal of Psychology. General Section*, *43*, 2-19.

Appendix

A linear mixed effect model with Task 2 RTs using the lmer() function of the R package lme4 yielded basically the same results as the ANOVA reported in the main text. Of the initial full model (d = determiner; s = sentence type; soa = SOA)

$$RT2 \sim d \cdot s \cdot soa + (1 + d \cdot s \cdot soa|subject) + (1 + d \cdot s \cdot soa|item),$$

the random effect structure was reduced to

$$RT2 \sim d \cdot s \cdot soa + (1 + d + s + soa + d{:}s + d{:}i + i{:}s||subject) + (1 + d||item),$$

to avoid singularities. Then, the significance pattern of fixed effects was assessed using Satterthwaite's method to estimate degrees of freedom. The results are summarized in Table A1 and the qualitative pattern resembles that of the standard ANOVA as reported in the main text. In particular, the effects of determiner and sentence type were significant, but neither of them interacted with SOA. Distributional assumptions were better complied with when RTs were log-transformed. The same analyses on log-transformed RTs yielded comparable results.

A generalized linear mixed effect model with a logit-link and the same random effect structure was then applied to Task 2 PEs using the glmer() function of the R package lme4 and the results are summarized in Table A1. Results resemble that of the ANOVA approach, except for the main effect of sentence type which was significant here as well.

Table A1. Inferential statistics from the linear mixed effect model analysis on Task 2 RT2s and the generalized linear mixed effect model on Task 2 PEs.

	Task 2 RTs			Task 2 PEs	
Effect	**|*t*|**	***dfs***	***p***	**|*z*|**	***p***
determiner	3.10	49.27	.003	4.99	<.001
sentence type	7.20	48.49	<.001	2.04	.041
soa	35.12	48.23	<.001	0.65	.517
determiner × sentence type	3.96	48.32	<.001	0.87	.386
determiner × soa	0.13	47.34	.899	0.60	.547
sentence type × soa	0.93	47.16	.357	0.22	.824
determiner × sentence type × soa	0.94	5759.19	.345	0.70	.487

10.3 Study 3

Official citation:

Schneider, C., Bade, N., & Janczyk, M. (2020). Is immediate processing of presupposition triggers automatic or capacity-limited? A combination of the PRP approach with a self-paced reading task. *Journal of Psycholinguistic Research, 49,* 247-273.

The final publication is available at https://www.springer.com/journal/10936

https://link.springer.com/article/10.1007/s10936-019-09686-3

Journal of Psycholinguistic Research
https://doi.org/10.1007/s10936-019-09686-3

Is Immediate Processing of Presupposition Triggers Automatic or Capacity-Limited? A Combination of the PRP Approach with a Self-Paced Reading Task

Cosima Schneider[1] · **Nadine Bade**[2] · **Markus Janczyk**[3]

Abstract
Informally speaking, presuppositions are meaning components which are part of the common ground for speakers in a conversation, that is, background information which is taken for granted by interlocutors. The current literature suggests an immediate processing of presuppositions, starting directly on the word triggering the presupposition. In the present paper, we focused on two presupposition triggers in German, the definite determiner *the* (German *der*) and the iterative particle *again* (German *wieder*). Experiment 1 replicates the immediate effects which were previously observed in a self-paced reading study. Experiment 2 then investigates whether this immediate processing of presuppositions is automatic or capacity-limited by employing the psychological refractory period approach and the locus of slack-logic, which have been successfully employed for this reason in various fields of cognitive psychology. The results argue against automatic processing, but rather suggest that the immediate processing of presuppositions is capacity-limited. This potentially helps specifying the nature of the involved processes; for example, a memory search for a potential referent.

Keywords Presuppositions · Experimental pragmatics · Dual-task · PRP

Introduction

Language and communication are ubiquitous in everyday life and speakers often communicate more than they actually say. How this additional meaning arises is an important question in the study of natural language meaning. Presuppositions are an example of meaning components that can be distinguished from the purely asserted meaning of an utterance, and have been a vital topic in the semantic and pragmatic literature of the last decades (see Beaver and Geurts 2012). While much of the previous work on presupposition processing

✉ Cosima Schneider
cosima.schneider@uni-tuebingen.de

1 Department of Psychology, University of Tübingen, Schleichstraße 4, 72076 Tübingen, Germany

2 Department of Cognitive Studies, École Normale Supérieure, Paris, France

3 Department of Psychology, University of Bremen, Bremen, Germany

Published online: 06 February 2020

focused on the influence of different contexts on the interpretation of presuppositions, the main goal of the present paper is to investigate at which stage of cognitive processing presuppositions unfold their impact.

Presuppositions and Their Immediate Processing

From a theoretical point of view, the term *presupposition* refers to background information which is taken for granted by speaker and listener. It differs from the *assertion* of a sentence, which is novel content and part of the main meaning of an utterance. Presuppositions are modeled as restrictions on what are appropriate contexts for the utterance (Heim 1991; Heim and Kratzer 1998; Stalnaker 1973), that is, propositions that must be entailed by the context in order for a sentence with a presupposition to be felicitously uttered and added to the common ground (Heim 1990). The context (set) or common ground is defined as the set of propositions believed to be true by all participants of a conversation. More formally speaking, a sentence p presupposes q if the use of p is inappropriate when q is not in the common ground (Stalnaker 2002). Under a semantic view, certain linguistic expressions trigger these appropriateness conditions and are therefore called *presupposition triggers*. In (1), for example, the word *again* triggers the presupposition that Anna has already scored before yesterday.

(1) Yesterday, Anna scored again.

As a result, a sentence as in (1) is predicted to only be appropriate (i.e., felicitous) in contexts which entail that Anna scored before. If the context does not entail this information, the sentence is predicted to be infelicitous. There is, however, a rescue strategy for sentences like (1) if the presupposition is not fulfilled. So-called *accommodation* describes the process of just assuming the presupposition to hold on the part of the speaker. It has been observed that accommodation is a highly context-dependent process (based on the probability of the truth of the presupposition in the given context; Heim 1992). For example, (1) might be surprising given that Anna never plays soccer. However, if she is known to be a very talented striker it is quite unsurprising. It has also been claimed that the availability of accommodation is dependent on the type of trigger and more difficult for triggers like *again* (see more discussion below).

Presuppositions are differentiated from asserted meaning and conversational implicatures, because they have different properties. For example, unlike assertions, presuppositions survive embedding under certain operators such as negation, conditionals, modals, or questions. The sentence in (2), for example, still presupposes that Anna scored again. However, it does not assert anymore that she scored yesterday.

(2) If Anna scored again yesterday, I'd be surprised.

As pointed out above, presuppositions are assumed to be encoded in a lexical trigger according to a semantic view, that is, they are associated with certain words (Frege 1892; Heim 1982; Russell 1905). This view is in line with the prediction that the trigger itself leads to awareness of the importance of context and could thus evoke immediate processing costs. However, there is an alternative theoretical perspective on presuppositions which takes a more pragmatic approach (Stalnaker 1973; Levinson 1983; Simons 2001). It assumes that presuppositions are not semantically encoded but are pragmatic, that is, they only play a role after the sentence's main meaning is computed and its integration into the context is considered. This "two-step" procedure means that the presupposition is not necessarily processed immediately, but only later at the end of the sentence. How presuppositions arise is still a highly debated issue in the literature ("the triggering problem"). It led to the debate whether presuppositions are needed as a separate concept or whether the issue is better understood in terms of what is at-issue or raises attention versus what is non-at-issue/in the background (Simons et al. 2011; Abrusán 2011; Tonhauser et al. 2018).[1] So far, there is a lot of evidence supporting the view that presuppositions are processed immediately (see below), which speaks against a two-step process. The data thus suggest that any processing model of presuppositions should contain the trigger itself as an important factor.

Experimental evidence for immediate processing of presuppositions comes from various methods. For example, Kirsten et al. (2014) investigated the processing of presuppositions while measuring event related potentials (ERPs) of the EEG in an experiment focusing on the presuppositions triggered by the definite determiner, compared to inferences arising from the indefinite determiner. Participants were presented with test sentences word-by-word on a computer screen and were asked comprehension questions at the end of the experiment. The data revealed ERP effects already on the trigger word. This led the authors to conclude that presupposition processing begins as soon as the presupposition trigger is encountered. Burkhardt's (2006) ERP study further supports the idea of early processing of presuppositions by revealing an N400 effect on the trigger position when the existence presupposition of the definite determiner was not given. The experiment varied the degree of availability of referents for definite determiner phrases by manipulating the context (*given*, *bridged*, and *new*). Definite noun phrases that were completely novel elicited N400 and P600 components compared to definite noun phrases whose referents were given in the context. In cases where the referent could easily be inferred (e.g., "the bus driver" in situations describing somebody entering a bus), the effect was weaker. In a follow-up study, Burkhardt (2007) manipulated the terms of inferential demands needed to form a relationship between the definite noun phrase and the information of the context sentence, which was previously presented. It was either necessary or inducible information. Drawing more demanding inferences resulted in larger P600 effects, whereas no N400 effects were observed when the context did not support the presupposition. Jouravlev et al. (2016) also examined ERPs, but focused on the PSP trigger *again* (in English). Participants read sentences in contexts that either supported the presupposition (e.g., "Jake had tipped a maid at the hotel once before. Today he tipped a maid at the hotel again…") or violated it (e.g., "Jake had never tipped a maid at the hotel before. Today he tipped a maid at the hotel again…"). The data analysis revealed the expected effects for semantic and syntactic violations (N440 and P600). Summing up, these results provide evidence for a rapid, on-line integration of presupposed content triggered by the adverb *again*. However, the observed pattern differs from the pattern reported for definite determiners.

Domaneschi et al. (2018) also investigated presupposition processing in different contextual conditions. To this end, they used contexts that satisfied the presupposition versus contexts that were neutral with regard to the truth of the presupposition (i.e., required accommodation), and compared two types of triggers, that is, definite descriptions and

[1] We are simplifying a bit here, as these accounts do not necessarily make these claims for all presupposition triggers but make further distinctions. As the direct comparison of these different classes of triggers is not the main focus of our paper, we will not go into the details of this part of the debate.

change-of-state verbs. The results also support the idea of immediate presupposition processing in the accommodation condition (a biphasic N400–P600 pattern at the point where the presupposition is known), but furthermore show that the two triggers differ in processing: for definite descriptions, a clear involvement of the N400 was observed, while for change-of-state verbs the costs of accommodation were associated with a more pronounced P600. The data support the idea that presupposition accommodation involves two steps: (1) search for a previous antecedent in the discourse, and in case of an unsuccessful search, (2) a second step of context repairment, namely an integration of the presupposed content into the discourse model.

In sum, these EEG studies provide evidence for an immediate processing of presuppositions, starting on the trigger itself. It is important to note that all of the studies presented focused on the influence of context, that is, they compared the processing cost of accommodation with the processing costs of a satisfied presupposition. In contrast, the present study focused on comparing a presupposition trigger with non-trigger words, and on the question whether processing the trigger is a capacity-limited process.

Other studies on presupposition processing used reading times. For example, Schwarz (2007) focused on the German additive particle and presupposition trigger *auch* (Engl. *too*) and reported longer reading times for clauses containing the trigger *auch* when the presupposition was not satisfied compared to when it was. Of particular importance for the present purposes is Experiment 1 of Tiemann et al. (2011). These authors also employed self-paced reading to investigate at which point in time processing of presuppositions takes place and included five different presupposition triggers (German *wieder*, Engl. *again*; *auch*, Engl. *also*; *aufhören*, Engl. *stop*; *wissen*, Engl. *know*; and definites in the shape of possessive noun phrases [*sein/ihr,* Engl. *his/her*]). In their experiment, they compared (1) sentences with a presupposition trigger, (2) grammatical sentences without a trigger, and (3) ungrammatical sentences without a trigger. The sentences were presented in contexts which did not explicitly verify the presupposition (i.e., they were neutral with regard to the presupposition). Overall, reading times at the positions of the trigger and the following word were longest in sentences with presupposition triggers, intermediate in grammatical sentences, and shortest in the ungrammatical sentences. These effects also indicate that a presupposition trigger is considered immediately upon encountering it. However, recent studies suggest that different types of presupposition triggers differ in processing (Abrusán 2011; Domaneschi et al. 2014; Domaneschi et al. 2018; Domaneschi and Di Paola 2018; Jouravlev et al. 2016; Tiemann et al. 2015). Against this background, it is unfortunate that Tiemann et al. (2011) did not analyze reading times for the different triggers separately. It thus remains unclear whether the results are similar for all triggers or just for a subset of them.

In sum, the current literature suggests an immediate processing of presuppositions, which starts directly on the trigger. The present study goes a step further by asking whether this immediate processing is automatic or capacity-limited. More precisely, we investigated this for two selected triggers, definite determiners and *again*, using a similar methodology as Tiemann et al. (2011, Exp. 1). The choice of triggers is partly motivated by the theoretical discussion in Kripke (2009), who argued that presuppositions triggered by *again* and *too* are especially hard to accommodate compared to definite determiners. The choice is also motivated by the classifications that were suggested to account for differences in processing. More specifically, Tiemann et al. (2015) suggested to categorize the triggers *again* and definite determiner in two different classes based on their different behavior. They proposed a maxim of interpretation which they called *Minimize Accommodation*: "Do not accommodate a presupposition unless missing accommodation will lead to uninterpretability of the assertion." According to this classification, Class 1 comprises triggers that are likely to be ignored in case of presupposition failure (e.g., particles like *again*, *too*, and *even*), because their presuppositions are not relevant to the assertion (and can thus be ignored given *Minimize Accommodation*). On the other hand, presuppositions of triggers in Class 2 must be accommodated according to this view, because otherwise the utterance cannot be interpreted (e.g., definite descriptions, factives, and change of state verbs), as these triggers do contribute to the assertion (see also Glanzberg 2005, for a similar distinction). Processing of presuppositions associated with the definite determiner and *again* should be different following this proposal: *again*, being a Class 1 trigger, does not contribute anything to the assertion of the sentence. That is, the sentence in (3) can be evaluated with regard to its truth conditional content (that Jenna went ice-skating) without knowing the presupposition. This is not the case for triggers belonging to Class 2 such as, for example, definite determiners. The truth of the sentence in (4) cannot be evaluated without the presupposition of existence and uniqueness being verified, that is, without knowing whether there is a sun and whether it is unique.

(3) Jenna went ice-skating, again.
(4) The sun is shining.

We therefore focus on these two triggers, which have been argued to belong to different categories. Focusing on only two triggers has the advantage that we will be able to increase the number of stimuli per participant to allow for meaningful separate analyses of the two triggers.

The Locus of Slack-Logic and an Example Application

To determine whether presupposition processing is automatic or a capacity-limited process, we will use the psychological refractory period (PRP) approach, a method that has been widely used in cognitive psychology with its origin in dual-task research. Of particular importance is the locus of slack-logic (Schweickert 1978) within a PRP experiment. We will introduce the general logic with an experimental example in the following, and will adapt this logic to a self-paced reading task.

In general, participants perform two independent tasks in each trial of a PRP experiment. The critical manipulation is the stimulus onset asynchrony (SOA), which is the time between the presentation of the Task 1 stimulus (S1) and the Task 2 stimulus (S2). With a short SOA, the two tasks overlap temporally, whereas there is no or only little temporal overlap with long SOAs. The typically observed result pattern is that the response time in Task 1 (RT1) does not depend on SOAs, but those in Task 2 (RT2) become longer the shorter the SOA—the PRP effect (Telford 1931). The most widely accepted explanation for this observation is the central bottleneck model (e.g., Pashler 1994; Welford 1952; see Fig. 1a for an illustration). A starting assumption of this model is that processing of a task is split into three stages: (a) a precentral stage, (b) a central stage, and (c) a postcentral stage. The precentral stage has most often been related to (early) perceptual processing and the postcentral stage to motor processing and execution. It is assumed that these two stages can run in parallel with all other stages of simultaneously processed tasks. The central stage has originally been related to response selection (Pashler 1994), but other processes seem to require this stage as well, for example, encoding into short-term memory (Jolicoeur and Dell'Acqua 1998), selection of working memory items (Janczyk 2017), or

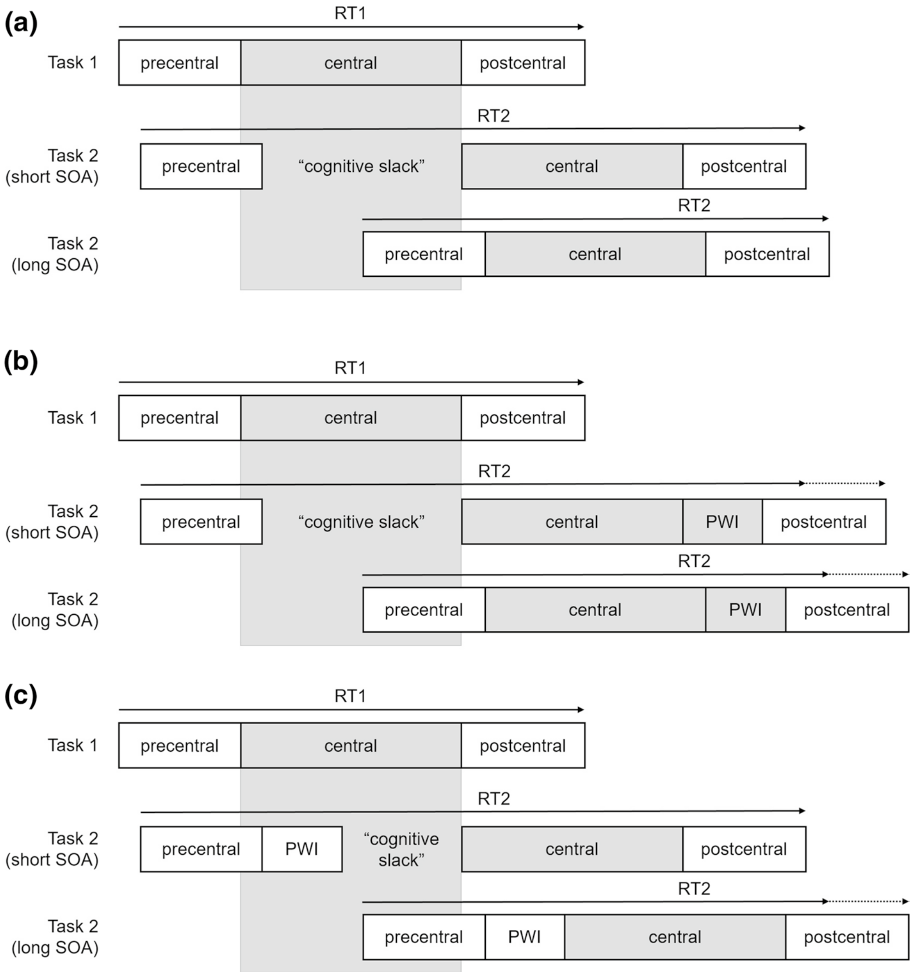


Fig. 1 Illustration of the central bottleneck model (**a**) and the predictions of the locus of slack-logic (**b**, **c**). (Note: *PWI* picture word interference)

anticipation of action effects (Wirth et al. 2015; see Janczyk and Kunde, under review). In contrast to the two other stages, the central stage is conceived as capacity-limited and can only be invoked by one task at a time, thereby constituting a bottleneck. With a short SOA, the central stage of Task 1 is not yet processed when the precentral stage of Task 2 has finished. Thus, central processing of Task 2 has to wait until the bottleneck is available again. This time of waiting is called the *cognitive slack* and is what leads to long RT2s with a short SOA. With a long SOA, in contrast, no cognitive slack occurs and Task 2 processing is not interrupted, resulting in short RT2s.

Importantly, this model can also be used to distinguish at which stage of processing a particular RT effect emerges (i.e., its "locus" in processing), and by implication then, whether this process is automatic or capacity-limited. We will explain this with a study by Piai et al. (2014) as an example, who investigated the locus of semantic interference in picture-word interference (PWI) experiments (see Abdel Rahman and Melinger 2009). Typically, participants are presented with pictured objects and distractor words and are instructed to name the picture while ignoring the distractor word. Naming latencies are shorter when picture and word match than when they do not. Piai et al. asked whether this semantic interference effect arises during the precentral stage, and thus is the result of parallel processing (Dell'Acqua et al. 2007), or during the capacity-limited central stage that was related to lexical selection (Schnur and Martin 2012). To illustrate, consider Piai et al.'s Experiment 1.[2] Task 1 was to give a manual response to a low- or high-pitched tone, and Task 2 was a vocal naming response to a picture combined with a distractor word. Pictures of the body parts leg, arm, and finger were combined with the corresponding word or a string of five Xs. In congruent trials, pictures and words matched, in incongruent trials, they did not match. In neutral trials, the pictures were presented with the five Xs. The SOA between the tone and the PWI stimulus was either 0 or 500 ms.

Two different predictions can be derived from the central bottleneck model. First, consider that the PWI effect results from processing during the capacity-limited central stage (see Fig. 1b). With a long SOA, Task 2 RTs are prolonged in incongruent compared to congruent trials (visualized by the gray box labeled PWI). Because with a short SOA the central stage can only start after the central stage of Task 1 has finished, the same PWI effect is expected in this case. In other words, the PWI effect is expected to combine additively with SOA. Second, consider that the PWI effect emerges from parallel processing that can run simultaneously with the central stage of Task 1 (see Fig. 1c). With the long SOA, the same prediction as for the previous case is made and the PWI effect should be observed. With a short SOA, in contrast, the processing leading to the PWI effects starts regardless of the central stage of Task 1, and any additional processing required in incongruent trials stretches into the cognitive slack. As a consequence, the PWI effect becomes invisible at the short SOA and SOA and PWI are expected to produce an (underadditive) interaction. The data clearly revealed an additive effect of SOA and PWI what suggests that the PWI effect requires central capacity and arises during (or after) the central stage. The results of further experiments in Piai et al. (2014) support this, because the additivity robustly replicated across these other experiments.

The Present Study: Is Processing of Presupposition Triggers Capacity-Limited or Automatic?

In the present study, we will use the PRP approach we just introduced to investigate the processing of presuppositions triggered by *again* and by definite determiners in more detail. The major question of our study is whether processing initiated when encountering a presupposition trigger is automatic or requires limited capacities. Experiment 1 was designed after Experiment 1 of Tiemann et al. (2011) with several goals. First, we aimed at replicating the observation of longer reading times for triggers compared with neutral or unacceptable sentences (Tiemann et al. 2011; see also Schwarz 2007, for the trigger *auch* compared to the neutral word *vorher* [Engl. *earlier*]). Second, because we needed to use a slightly modified presentation method of the words in the self-paced reading task to apply the PRP setup and the locus of slack-logic in Experiment 2, we already adopted

[2] In this experiment, the authors actually compared performance in a PWI task with performance in a color word Stroop task. For simplicity, we here only consider the PWI task.

this method in Experiment 1 to ensure that the longer reading times for triggers are also observed under these conditions. Third, based on the acceptability ratings of sentences collected in Experiment 1, we selected those items that fit best for use in the subsequent experiment. In Experiment 2, we then adapted the PRP approach to the reading task by adding a tone discrimination task and presenting the trigger (or the corresponding word at this position) after a variable SOA following the tone. To ensure that participants interpreted the sentences in the intended way, we again included the rating after each trial and asked comprehension questions at the end of the experiment. We would like to stress at this point that conclusions about differences between the triggers can only be made if the qualitative pattern we observe is different. Numerical differences, even if substantiated by significant main effects, do not necessarily mean that the underlying processes are different. For example, the processes may simply require more time because they are more difficult in one condition.

Experiment 1

Experiment 1 uses a self-paced reading task to investigate and establish the reading times for several regions of interest (i.e., the presupposition trigger, the word following the presupposition trigger, the final word, and the total reading time) separately for two particular presupposition triggers, namely determiners and the German word *wieder* (Engl. *again*). Additionally, this experiment prepared the subsequent Experiment 2, which focuses on the main question of this paper. To this end, participants rated acceptability of sentences against the presented context after each trial. On the basis of these data, we selected the sentences for the following experiment. Furthermore, Experiment 2 required the simultaneous presentation of all words preceding the presupposition trigger or the corresponding word on the trigger position to apply the locus of slack-logic. Thus, we already used this procedure in Experiment 1 to determine whether or not we still observe an effect of the presupposition trigger in reading times.

Although this experiment is closely designed after Experiment 1 of Tiemann et al. (2011), we used only two triggers as opposed to the five different triggers used by Tiemann et al. This allowed us to increase the number of times each trigger was presented in the experiment. Following Tiemann et al., we will first visualize reading times averaged for both triggers, but—if warranted—this is followed-up by analyses of both triggers separately. By and large, the expectation was to replicate the results obtained by Tiemann et al. despite the changes in the presentation procedure and to identify possible differences between the two triggers belonging to different categories.

Method

Participants

Forty-eight native speakers of German (35 female, 13 male; mean age = 24.4 years) participated in this experiment. They were recruited from the participant pool at the University of Tübingen (Germany), were naïve regarding the hypotheses of this experiment, and signed informed consent prior to data collection. Participants received 8€ or course credit for their participation.

Apparatus and Stimuli

Stimulus presentation and response collection were controlled by a standard PC connected to a 17-in. CRT monitor. Responses in the reading task were given on an external response key which was located to the right of the participants and was operated with the right index-finger. Ratings of the sentences were provided via the number keys 1–4 on a standard QWERTZ keyboard ranging from very unnatural (1) to very natural (4).

All stimuli were presented in white font on a black background. Context sentences were presented in full length in the upper half of the screen. The letters of the test sentences' words were first substituted by underscores as placeholders. All words preceding the presupposition trigger or the corresponding word on this position were presented simultaneously; all subsequent words were presented one-by-one (see below, section "Task and Procedure" for more information). Once a new word was presented, the previous word disappeared and was again substituted with the underscores (see Fig. 2).

We included two types of presupposition triggers in this experiment, namely the German definite determiner *der* (Engl. *the*) and the German iterative particle *wieder* (Engl. *again*). For each trigger, we created 52 sets of experimental sentences, thus 104 sets in total. Each set consisted of a context sentence and three test sentences. The context sentences merely introduced the protagonists, but were kept as neutral as possible with regard to the truth of the presupposition. They were designed so that they made the acceptable test sentence appropriate, the trigger sentence somewhat degraded due to the presupposition being neither true nor false in the context, and the unacceptable sentence inappropriate [see (5) and (7)]. The test sentences contained either a presupposition trigger [(6a) and (8a)], a neutral word [(6b) and (8b)], or a semantically unacceptable word [(6c) and (8c)]. The neutral/unacceptable words replaced the trigger word in the respective conditions and kept the sentence semantically acceptable or made it semantically unacceptable. In total, 312 trials resulted.

Example item *again*

(5) Kontext: Monika ist mit ihren Freunden unterwegs.
Context: Monika is with her friends out.

(6) Test sentences:

(a) <u>Monika läuft</u> **wieder** Schlittschuh und lacht. (trigger)
<u>Monika does</u> **again** ice-skating and smiles.

(b) <u>Monika läuft</u> **heute** Schlittschuh und lacht. (neutral)
<u>Monika does</u> **today** ice-skating and smiles.

(c) <u>Monika läuft</u> **freundlich** Schlittschuh und lacht. (unacceptable)
<u>Monika does</u> **friendly** ice-skating and smiles.

Example item *determiner*

(7) Kontext: Marie sonnt sich heute im Garten.
Context: Marie suns herself today in (the) garden.

(8) Test sentences:

(a) Marie liegt auf **der** Liege und trinkt Wasser. (trigger)
Marie lies on **the** lounger and drinks water.

(b) Marie liegt auf **einer** Liege und trinkt Wasser. (neutral)
Marie lies on **a** lounger and drinks water.

(c) Marie liegt auf **jeder** Liege und trinkt Wasser. (unacceptable)
Marie lies on **every** lounger and drinks water.

When creating context and test sentences, we pursued the same goals as Tiemann et al. (2011) did. Most importantly, we made the sentences as neutral as possible with regard to the presupposition, that is, they did not explicitly verify or falsify it. At the same time, we made the events described plausible in the given setting so that the "neutral" test condition would be completely acceptable, the trigger sentence somewhat acceptable (requiring accommodation, however), and the unacceptable sentence the most unacceptable (as it was ill-formed irrespective of plausibility in the context).

Task and Procedure

Each trial started with the complete context sentence, horizontally centered in the upper part of the computer screen (see Fig. 2 for an illustration of the following). After participants read the sentence, they were to press the response button to request the test sentence. The test sentence was presented in a self-paced reading manner. This allows readers to use the response button presses to control the exposure duration for each section of the sentence they read. The test sentence was divided into a segment preceding the trigger word or the corresponding word on this position [the underlined part in Examples (6) and (8)], in which all words were presented simultaneously, and a section following it. Since simultaneous presentation applied to all sentence types, it was up to then equally likely for a participant to be confronted with a trigger sentence, a neutral sentence, or an unacceptable sentence. The following words, that is the presupposition trigger itself, the neutral word, or the unacceptable word [printed in bold font in Examples (6) and (8)], and all subsequent words were presented word-by-word upon response key presses. Reading times were measured from word/segment onset until the response key was pressed. After the test sentence was read, participants rated the acceptability of the test sentence within the given context.

Participants started with reading written instructions. This was followed by a short practice block with two sets of each trigger in all three conditions, thus 12 trials in total. The

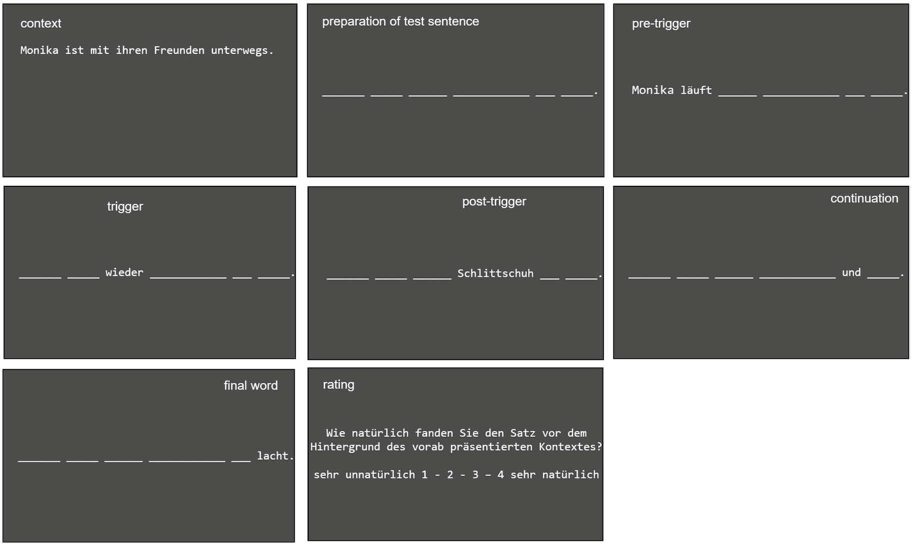


Fig. 2 Illustration of the task used in Experiment 1 (see text for more information). (Note that the words appearing in the upper part ("context", "preparation of test sentence", …) did not actually appear during the experiment but were added here for clarity)

order of these practice trials was determined randomly, but was the same for all participants. Then, the 300 test trials were administered in three blocks of 100 trials each. The order of presentation was random, with the restriction that sentences of the same item did not appear in different conditions directly in succession. All participants were tested individually in a single session of about 60 minutes. This is another slight change compared to the original study: Tiemann et al. (2011) tested participants in three separate sessions to avoid that they saw the same item in different conditions within one session. As we increased the number of stimuli though, we did not expect recognition effects during one session.

Design and Analyses

The independent variables of interest were (1) sentence type (trigger vs. neutral vs. unacceptable) and (2) trigger type (determiner vs. *again*). Mean acceptability ratings were submitted to a 3×2 Analysis of Variance (ANOVA) with sentence type and trigger type as repeated-measures. Reading times were calculated per letter (see Tiemann et al. 2011) for the following regions: (1) the word(s) preceding the trigger position (pre-trigger), (2) the presupposition trigger or the corresponding word on this position (trigger), (3) the word following the trigger position (post-trigger), the final word (final word), and the reading time of the whole sentence (total). Trials in which one reading time deviated more than 2.5 standard deviations from the respective design cell (calculated separately for each participant) were excluded as outliers (11.01% of the trials). Mean reading times for each region were submitted to the same ANOVA as acceptability ratings were. When the interaction of trigger type × sentence type was significant, we ran separate ANOVAs for both triggers with sentence type as a repeated-measure. A significant main effect in this analysis was followed up by paired *t* tests. In case of violations of the sphericity assumption, uncorrected degrees of freedom are reported, but the corresponding ε-estimate is provided. Effect sizes for *t* tests were calculated as $d = \frac{t}{\sqrt{n}}$ with $n = 48$.

Results

Acceptability Rating

Results of the acceptability rating are visualized in Fig. 3a. Unacceptable sentences were rated worst and trigger and neutral sentences were rated much more appropriate. Descriptively, for the determiner condition, ratings for neutral sentences were slightly worse than for trigger sentences, whereas for the trigger *again*, neutral sentences were rated best. The ANOVA revealed a main effect of sentences type, $F(2,94) = 330.60$, $p < .001$, $\eta_p^2 = .88$, $\varepsilon = .55$, and of trigger type, $F(1,47) = 4.78$, $p = .034$, $\eta_p^2 = .09$. The interaction was significant as well, $F(2,94) = 6.08$, $p = .007$, $\eta_p^2 = .11$, $\varepsilon = .77$, and we therefore analyzed the two triggers separately.

For the determiner condition, the ANOVA revealed a main effect of sentence type $F(2,94) = 302.87$, $p < .001$, $\eta_p^2 = .87$, $\varepsilon = .59$. Significant differences were obtained between all sentence types, trigger versus neutral: $t(47) = 3.10$, $p = .003$, $d = 0.45$; unacceptable versus trigger: $t(47) = 17.89$, $p < .001$, $d = 2.58$; unacceptable versus neutral: $t(47) = 17.89$, $p < .001$, $d = 2.58$. For the trigger *again*, the main effect of sentence type was significant as well, $F(2,94) = 213.66$, $p < .001$, $\eta_p^2 = .82$, $\varepsilon = .61$, and the *t* tests revealed significant differences between all sentence types, trigger versus neutral: $t(47) = 4.58$, $p < .001$, $d = 0.66$; trigger versus unacceptable: $t(47) = 14.62$, $p < .001$, $d = 2.11$; neutral versus unacceptable: $t(47) = 15.53$, $p < .001$, $d = 2.24$.

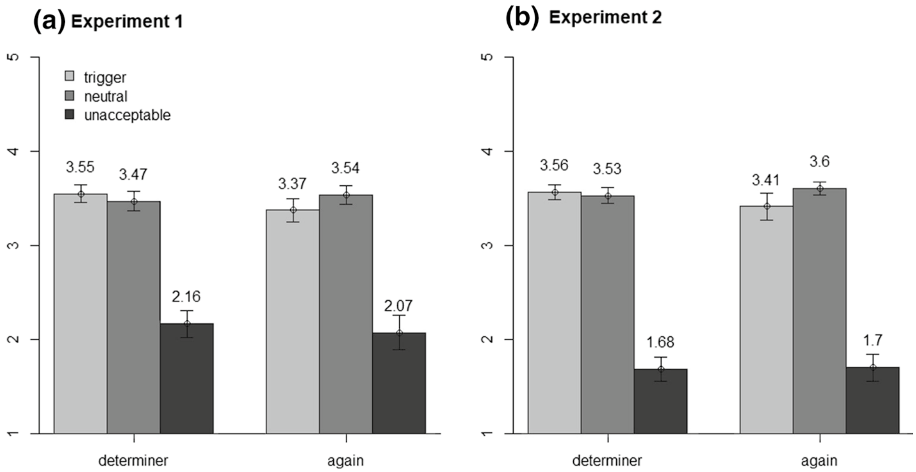


Fig. 3 Acceptability ratings in Experiments 1 and 2 as a function of sentence type and trigger type. Error bars are 95% confidence intervals for the means

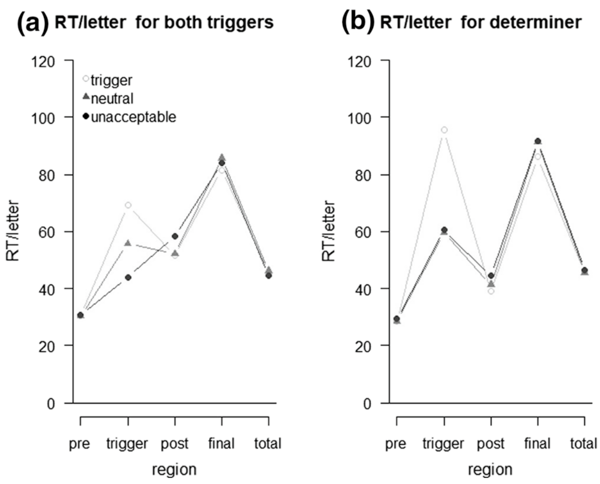


Fig. 4 Reading times (RT; in milliseconds) per letter of Experiment 1 analyzed across triggers in (**a**), and separately for the two triggers determiner (**b**) and *again* (**c**) for the regions pre-trigger (pre), trigger, post-trigger (post), final word (final), and total as a function of sentence type

Reading Times

Reading times per letter across both triggers are visualized in Fig. 4a, and separately for the determiner and *again* in Fig. 4b and c, respectively. All inferential statistics are summarized in Table 1. The ANOVA revealed significant differences between the two trigger types for all analyzed positions, perhaps pointing to differences in how the two triggers are processed.

For the trigger position, the interaction was significant and differences in reading times were observed for both trigger conditions, though in different directions. For the determiner, trigger sentences had the longest reading times, while those for neutral and unacceptable sentences did not differ. In contrast, for *again*, reading times were longest for neutral sentences, intermediate for trigger sentences, and shortest for unacceptable sentences.

Also for the post-trigger position, the interaction was significant and differences in reading times were observed for both triggers. For the determiner, differences were small in size, but reading times were longest for unacceptable sentences, intermediate for neutral sentences, and shortest for trigger sentences. For *again*, reading times were longest for unacceptable sentences, but similar for trigger and neutral sentences.

No differences in reading times between the sentence types were obtained for the final word. When considering the total reading time though, reading times depended on sentence type only for *again*, and were longest for neutral sentences, intermediate for trigger sentences, and shortest for unacceptable sentences.

Discussion

Experiment 1 was largely built on Experiment 1 of Tiemann et al. (2011), however, we focused on the definite determiner and *again* to allow for separate analyses of reading times if warranted. The rating data replicate the results of Tiemann et al. in general, with minor

Table 1 Inferential statistics for Experiment 1

	Pre-trigger	Trigger	Post-trigger	Final word	Total
Trigger type	$F(1,47)=82.37$, $p<.001$, $\eta_p^2=.64$	$F(1,47)=628.80$, $p<.001$, $\eta_p^2=.93$	$F(1,47)=472.96$, $p<.001$, $\eta_p^2=.91$	$F(1,47)=39.87$, $p<.001$, $\eta_p^2=.46$	$F(1,47)=6.69$, $p=.013$, $\eta_p^2=.12$
Sentence type	$F(2,94)=0.86$, $p=.426$, $\eta_p^2=.02$	$F(2,94)=544.50$, $p<.001$, $\eta_p^2=.92$, $\varepsilon=.73$	$F(2,94)=25.23$, $p<.001$, $\eta_p^2=.35$, $\varepsilon=.60$	$F(2,94)=0.73$, $p=.439$, $\eta_p^2=.02$, $\varepsilon=.69$	$F(2,94)=3.01$, $p=.073$, $\eta_p^2=.06$, $\varepsilon=.71$
Interaction	$F(2,94)=2.99$, $p=.055$, $\eta_p^2=.06$	$F(2,94)=409.19$, $p<.001$, $\eta_p^2=.90$, $\varepsilon=.88$	$F(2,94)=10.67$, $p<.001$, $\eta_p^2=.19$, $\varepsilon=.87$	$F(2,94)=1.11$, $p=.334$, $\eta_p^2=.02$	$F(2,94)=22.54$, $p<.001$, $\eta_p^2=.32$, $\varepsilon=.87$
Only determiner					
Sentence type		$F(2,94)=466.72$, $p<.001$, $\eta_p^2=.91$, $\varepsilon=.86$	$F(2,94)=8.30$, $p=.003$, $\eta_p^2=.15$, $\varepsilon=.63$		$F(2,94)=0.58$, $p=.530$, $\eta_p^2=.01$, $\varepsilon=.83$
Trigger versus neutral		$t(47)=25.10$, $p<.001$, $d=3.62$	$t(47)=3.46$, $p=.001$, $d=0.50$		
Trigger versus unacceptable		$t(47)=23.26$, $p<.001$, $d=3.36$	$t(47)=3.27$, $p=.002$, $d=0.47$		
Neutral versus unacceptable		$t(47)=0.81$, $p=.422$, $d=0.12$	$t(47)=2.09$, $p=.042$, $d=0.30$		
Only *again*					
Sentence type		$F(2,94)=502.61$, $p<.001$, $\eta_p^2=.91$, $\varepsilon=.76$	$F(2,94)=37.73$, $p<.001$, $\eta_p^2=.45$, $\varepsilon=.73$		$F(2,94)=15.37$, $p<.001$, $\eta_p^2=.25$, $\varepsilon=.71$
Trigger versus neutral		$t(47)=13.60$, $p<.001$, $d=1.96$	$t(47)=1.58$, $p=.121$, $d=0.23$		$t(47)=3.53$, $p=.001$, $d=0.51$
Trigger versus unacceptable		$t(47)=23.01$, $p<.001$, $d=3.32$	$t(47)=6.04$, $p<.001$, $d=0.87$		$t(47)=2.97$, $p=.005$, $d=0.43$
Neutral versus unacceptable		$t(47)=25.06$, $p<.001$, $d=3.62$	$t(47)=7.07$, $p<.001$, $d=1.02$		$t(47)=4.73$, $p<.001$, $d=0.68$

The first rows are the statistics for the 3×2 ANOVA for each region. In case of a significant interaction, separate ANOVAs with sentence type as a repeated-measure were run. If these were significant, the three sentence types were compared with paired *t* tests exceptions: Unacceptable sentences were rated worst, and for the trigger *again*, neutral sentences were rated slightly better than trigger sentences. In contrast to Tiemann et al.'s study, trigger sentences were rated better than neutral sentences for the definite determiner. In the original study this was reversed although it is unclear from the report whether the contrasts between sentence types were significant. Overall, this supports the original idea of Tiemann et al. that using presuppositions in neutral contexts is not as unacceptable as using grammatically deviant structures. As a result, successful context integration (i.e., accommodation of the presupposition) should be distinguished from semantic violations.

That context integration did play a role, that is, that participants accommodated the presupposition, is supported by the ratings for the trigger condition, which are unexpectedly quite high, and higher than in the original study. Although the presupposition was not actually mentioned in the context, participants easily accepted the sentences. This suggest that a process of accommodation took place, which was facilitated by the contexts we used. The deviation from Tiemann et al.'s results can be explained by assuming that the contexts used in the present study made accommodation more likely. The observed difference between *again* and the determiner is rooted in the fact that the presuppositions of determiners in general seem to be easier to accommodate (Tiemann et al. 2015).[3] Based on the ratings, we selected the 32 items that fit our requirements best for use in Experiment 2, namely those sentences that revealed the general pattern we expected most clearly (ungrammatical sentences are worse than trigger sentences which are [slighlty] worse than acceptable sentences).

Reading time results are largely in line with Tiemann et al.'s (2011) observations, but also extend them in an important way. Most importantly, we were able to replicate immediate effects on the trigger and the word following the trigger, with a descriptive pattern very similar to the original study. These results speak for an immediate processing of the presupposition trigger. However, one purpose of the present study was to analyze both triggers separately. While for both trigger types reading times for the trigger positions were longer for trigger than for unacceptable sentences, neutral sentences had the longest reading times for the trigger *again*, but for the determiner, they were similar to those of unacceptable sentences. The long reading times for the neutral condition for the trigger *again* might be due to the unexpected appearance of the word *heute* (Engl. *today*) in this position. It sounds more natural to place the word *heute* at the beginning of the sentence in German. This unexpected word order might have caused the long reading times.

In sum, Experiment 1 replicated effects already on the trigger position for both trigger types, despite our change of presenting all pre-trigger position words simultaneously and testing all items in one session.

Experiment 2

By and large, Experiment 1 replicated and extended the results obtained by Tiemann et al. (2011). Based on this, Experiment 2 embeds the self-paced reading task within a PRP experiment to apply the locus of slack-logic. The goal is to evaluate whether the processing

[3] We would like to point out though that we can only draw weak conclusions about accommodability based on our data, as we did not explicitly control for factors that have been shown to influence the availability of accommodation, for example, plausibility, bridging etc. It is important to note, however, that this is not the main point of the present study, but rather the comparison of presuppositional sentences to non-presuppositional ones, both grammatical and ungrammatical, to test whether processing presuppositions is capacity limited.

initiated by a presupposition trigger is (a) automatic and running in parallel with other tasks or is (b) capacity limited with a locus within the central stage of processing. Thus, a binary tone discrimination was added to the self-paced reading task. More precisely, a tone was played after participants read all pre-trigger position words and participants were to respond with a key-press with their left hand to the pitch of the tone. After a variable SOA, the word on the trigger position appeared, and participants proceeded through the remaining sentence in a similar way as in Experiment 1. In terms of the PRP logic (see Introduction), the tone discrimination task can be considered as Task 1, and reading the word on the trigger position would be Task 2.

Because the locus of slack-logic can—in the present setup—only be applied to the word on the trigger position, the predictions for Experiment 2 focus on this position.[4] We illustrate the predictions for the comparison between trigger and unacceptable sentences in Fig. 5, with the former sentences having resulted in longer reading times for both triggers in Experiment 1. Regardless of whether trigger processing is capacity-limited (Fig. 5a) or automatic (Fig. 5b), differences in reading times for the trigger position are expected with a long SOA. Ideally, the pattern observed there should be the same as already obtained in Experiment 1. Different predictions, however, can be made for the situation with a short SOA. If processing at the trigger position does require central capacity (Fig. 5a), it cannot be initiated before the central stage of Task 1 has finished. In this case, the same differences as with the long SOA are observed and—statistically—sentence type and SOA should combine additively. If, in contrast, this processing is automatic and runs in parallel to the central stage of Task 1, all differences become absorbed into the cognitive slack and should become unobservable with the short SOA (Fig. 5b). Statistically, sentence type and SOA should yield an (underadditive) interaction. In any case, reading times are expected to be longer with a short than with a long SOA, that is, a PRP effect, because some central processing can be assumed anyway, for example, response selection required for pressing the response key (see also Janczyk 2017, for an example).

Method

Participants

The intended sample size in this experiment was $n=48$. Data were collected from 51 native speakers of German from the Tübingen (Germany) area, of which three participants were excluded because of 30% or more errors in the comprehension questions at the end of the experiment (final sample: mean age = 24.2 years, 39 females, 9 males). Participants signed informed consent prior to data collection and were paid 8€ or received course credit for participation.

Apparatus and Stimuli

The same general setup as in Experiment 1 was used. Due to the addition of the auditory (tone) discrimination task, two additional response keys were placed to the left of the

[4] In principle, of course, the PRP and locus of slack-methods can be applied to other positions as well. Yet, one has to conduct a separate experiment, because with the locus of slack-logic it is only possible to investigate one position at a time. Thus, a similar investigation, for example, for the point where the presupposition becomes completely known cannot be done within the same experiment and must be postponed to future research.

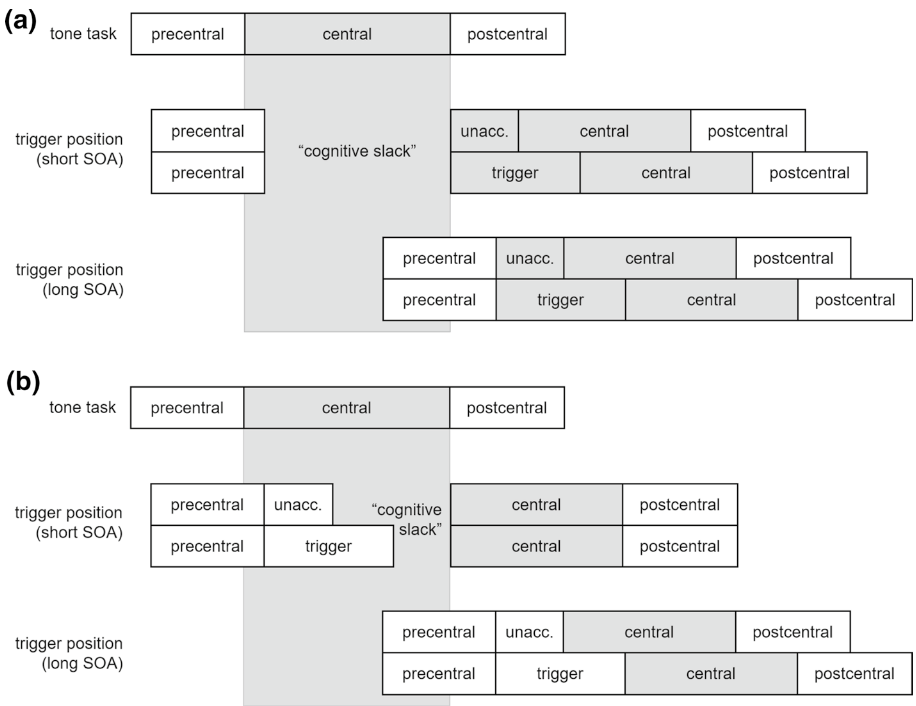


Fig. 5 Predictions for reading times at the trigger position for the comparison trigger versus unacceptable (unacc.) sentences. **a** Processing at the trigger position requires central capacity and can thus only start once the central stage of the (preceding) tone task has finished. **b** Processing at the trigger position can run in parallel to other capacity-limited stages and with a short SOA it extends into the cognitive slack. (*SOA* stimulus onset asynchrony)

participants which were operated with the left index- and middle-finger. Stimuli in this task were 300 and 900 Hz sinusoidal tones of 50 ms length presented via headphones.

Based on the results of Experiment 1, we selected the 32 sets of experimental items per trigger (out of the 50 experimental sets used in Experiment 1) that best fit our expectations regarding the ratings (see "Appendix" for more details on the selection).

Task and Procedure

While the general procedure of the self-paced reading task and the acceptability rating was similar to Experiment 1 (see Fig. 6), several changes were required to integrate the auditory discrimination task. To minimize exclusion of trials due to errors in this task, participants started with 40 practice trials of only the auditory discrimination task. Each of these trials was initiated by written instructions centered on the screen that asked the participant to press the right response key to start a trial (to mimic the procedure required in the main experiment). After 100 ms, the 300 or 900 Hz tone was played (each 20 times in a random order) and participants were to respond with a key press of the left hand. In case of errors, respective error feedback was presented on the screen (1000 ms) and the next trial started after an inter-trial interval of 1000 ms.

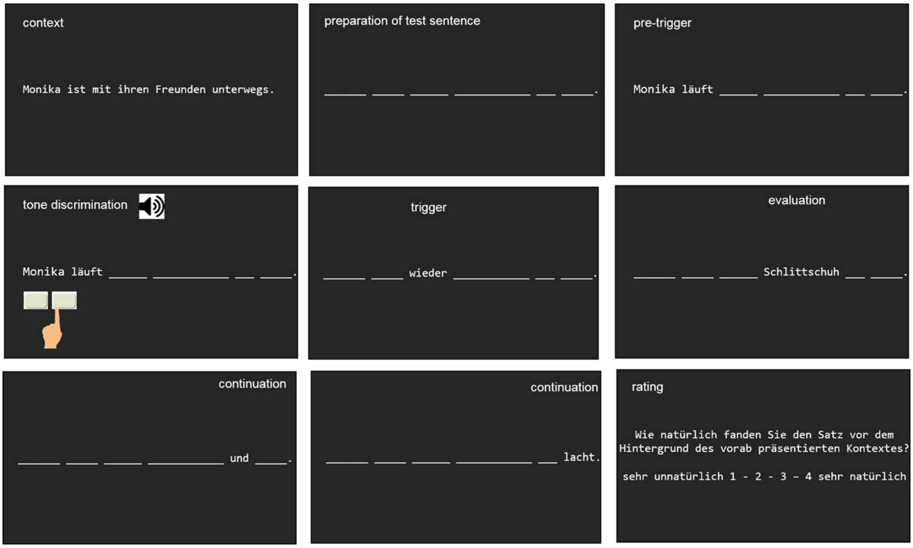


Fig. 6 Illustration of the task used in Experiment 2 (see text for more information). (Note that the words appearing in the upper part ("context", "preparation of test sentence", …) and the pictures illustrating the tone discrimination task did not actually appear during the experiment but were added here for clarity)

Then the main experiment began and, similarly to Experiment 1, participants read a context sentence at the beginning of each trial. Pressing the right response key made the context sentence disappear and the underscores appeared as placeholders. With the next key press, all pre-trigger position words appeared. Then, following the next press of the right response key, the 300 or 900 Hz tone was played and following an SOA of 100 or 1200 ms, the next word (on the trigger position) occurred. Participants were to first respond to the tone and then to continue reading. If the response to the tone was correct and no further errors occurred (wrong response order, no response within 5000 ms), the trial was continued until the end of the sentence. As in Experiment 1, this procedure was followed by an acceptability rating task, but in case of errors, the trial was aborted and respective error feedback was presented on the screen (1000 ms).

Participants began with 48 (unanalyzed) practice trials. The 32 (sets) × 2 (trigger types) × 3 (sentence type) = 192 test trials were divided into three test blocks of 64 test sentences each, with a randomized presentation order. The two SOAs were orthogonally distributed across the six combinations of trigger types and sentence types. The stimulus–response mapping of the auditory discrimination task was counterbalanced across participants.

Design and Analyses

For all analyses (except on error rates in the discrimination task), trials with an erroneous response in the auditory discrimination task were excluded first (those trials were aborted during the experiment and not continued). Acceptability ratings were analyzed as for Experiment 1. Mean discrimination response times, error rates in the discrimination task, and reading times (per letter) at the trigger position were submitted to a 2 × 3 × 2 ANOVA with SOA (100 vs. 1200 ms), sentence type (trigger vs. neutral vs. unacceptable), and trigger type (determiner vs. again) as repeated-measures.

Trials with response times or reading times deviating more than 2.5 standard deviations from the mean of the respective design cell (calculated separately for each participant) were excluded as outliers.

Results

First of all, we excluded 2.7% trials with unspecific errors (too slow response, responding to the reading task before responding to the discrimination task, no response within the time limit).

Acceptability Rating

Results of the acceptability rating are visualized in Fig. 3b. For the determiner condition, trigger and neutral sentences were rated almost equally well, while unacceptable sentences were rated worst. For *again*, neutral sentences were rated better than trigger sentences, while unacceptable sentences were also rated worst. The ANOVA revealed a main effect of sentence type, $F(2,94)=551.27$, $p<.001$, $\eta_p^2=.92$, $\varepsilon=.63$, but not of trigger type, $F(1,47)=0.29$, $p=.595$, $\eta_p^2=.01$. The interaction was also significant, $F(2,94)=4.27$, $p=.025$, $\eta_p^2=.08$, $\varepsilon=.79$, and we thus analyzed the two triggers separately.

For the determiner, we observed a main effect of sentence type, $F(2,94)=478.90$, $p<.001$, $\eta_p^2=.91$, $\varepsilon=.59$. There was no significant difference between trigger and neutral sentences, $t(47)=1.40$, $p=.167$, $d=0.20$, but ratings for unacceptable sentences differed significantly from both the trigger sentences, $t(47)=22.65$, $p<.001$, $d=3.27$, and the neutral sentences, $t(47)=22.30$, $p<.001$, $d=3.22$. For *again*, the main effect of sentence type was also significant $F(2,94)=331.67$, $p<.001$, $\eta_p^2=.88$., $\varepsilon=.88$, and all differences were significant, trigger versus neutral: $t(47)=2.88$, $p=.006$, $d=0.42$; trigger versus unacceptable: $t(47)=18.49$, $p<.001$, $d=2.67$; neutral versus unacceptable: $t(47)=22.94$, $p<.001$, $d=3.31$.

Auditory Discrimination Task

Response times and error percentages are summarized in Table 2 (2.55% outliers). Participants responded more slowly with a short compared to a long SOA, $F(1,47)=47.57$, $p<.001$, $\eta_p^2=.50$, and overall descriptively slightly shorter in the *again* condition, $F(1,47)=2.46$, $p=.123$, $\eta_p^2=.05$. There was also a main effect of sentence type, with slowest responses for unacceptable sentences, intermediate for neutral sentences, and fastest responses for the trigger sentences, $F(2,94)=5.02$, $p=.013$, $\eta_p^2=.10$, $\varepsilon=.84$. This effect was much larger at the short than at the long SOA, thus an overadditive interaction, $F(2,94)=9.87$, $p=.001$, $\eta_p^2=.17$, $\varepsilon=.77$, and for *again* compared to the determiner, $F(2,94)=4.20$, $p=.018$, $\eta_p^2=.08$. There was no significant interaction of SOA and trigger type, $F(1,47)=0.40$, $p=.532$, $\eta_p^2=.01$. Because the three-way interaction was significant, however, $F(2,94)=3.72$, $p=.028$, $\eta_p^2=.07$, separate 3×2 ANOVAs were run for each trigger type.

For the determiner, only the main effect of SOA was significant, $F(1,47)=42.72$, $p<.001$, $\eta_p^2=.48$. Neither the main effect of sentence type, $F(2,94)=0.33$, $p=.722$,

Table 2 Mean response times (in milliseconds)|error percentages for the auditory discrimination task as a function of sentence type, trigger type, and stimulus onset asynchrony (SOA)

Sentence type	Trigger type			
	Determiner		Again	
	SOA (ms)		SOA (ms)	
	100	1200	100	1200
Trigger	823|3.63	763|1.70	801|2.49	734|1.83
Acceptable	843|1.98	752|2.25	810|2.41	751|2.15
Unacceptable	860|3.29	743|1.97	891|1.60	732|2.92

$\eta_p^2=.01$, nor the interaction were significant, $F(2,94)=3.04$, $p=.059$, $\eta_p^2=.06$, $\varepsilon=.89$. For *again*, both main effects were significant, SOA: $F(1,47)=40.15$, $p<.001$, $\eta_p^2=.46$; sentence type: $F(2,94)=9.28$, $p<.001$, $\eta_p^2=.16$, as was the interaction, $F(2,94)=12.53$, $p<.001$, $\eta_p^2=.21$, $\varepsilon=.84$.

For error rates, no effect reached significance, all $Fs\leq 3.06$, all $ps\geq .087$.

Self-Paced Reading Task

Mean reading times (per letter) for the trigger position across both trigger types are visualized in Fig. 7a (2.77% outliers). Sentence type influenced reading times significantly, $F(2,94)=471.53$, $p<.001$, $\eta_p^2=.91$, $\varepsilon=.76$. Furthermore, we observed a main effect of SOA, $F(1,47)=1539.54$, $p<.001$, $\eta_p^2=.97$, with shorter reading times with a long SOA, and a main effect of trigger type, $F(1,47)=495.16$, $p<.001$, $\eta_p^2=.91$. Sentence type and SOA also interacted, $F(2,94)=231.57$, $p<.001$, $\eta_p^2=.83$, but the effect of sentence type was *larger* with a short than with a long SOA, that is, the interaction was *over*additive.

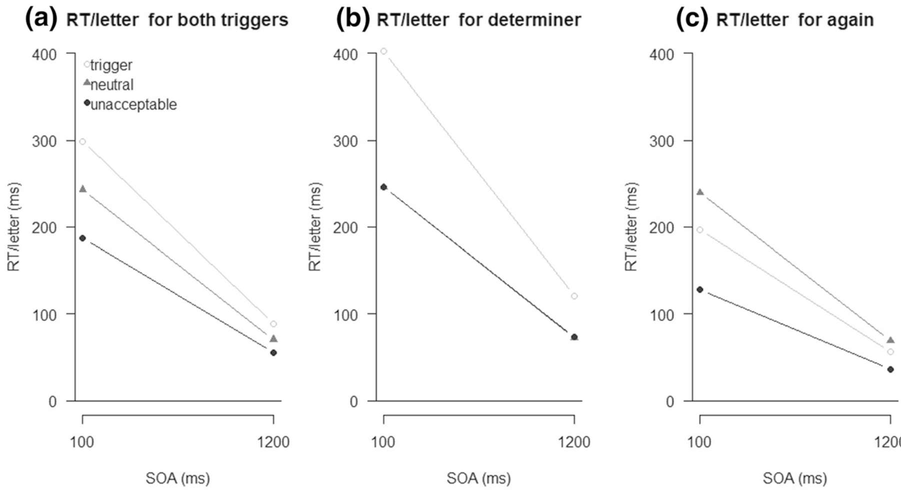


Fig. 7 Reading times (RT) for the trigger position as a function of sentence type and stimulus onset asynchrony (SOA) in Experiment 2 analyzed across triggers in (**a**), and separately for the two triggers determiner (**b**) and *again* (**c**)

Sentence type and trigger type also interacted, $F(2,94) = 376.94$, $p < .001$, $\eta_p^2 = .89$, $\varepsilon = .67$, and the interaction of SOA and trigger type was significant as well, $F(1,47) = 717.47$, $p < .001$, $\eta_p^2 = .94$. Finally, the three-way interaction was also significant, $F(2,94) = 175.37$, $p < .001$, $\eta_p^2 = .79$, $\varepsilon = .84$, and we thus analyzed both trigger types separately.

For the trigger determiner (Fig. 7b), the main effect of sentence type was significant, $F(2,94) = 391.44$, $p < .001$, $\eta_p^2 = .89$, $\varepsilon = .71$, as was the main effect of SOA, $F(1,47) = 1590.31$, $p < .001$, $\eta_p^2 = .97$. Further, we observed a significant interaction of sentence type and SOA, $F(2,94) = 186.88$, $p < .001$, $\eta_p^2 = .80$, $\varepsilon = .82$. A very similar picture was obtained for the trigger *again* (Fig. 7c). The main effect of sentence type was significant, $F(2,94) = 542.56$, $p < .001$, $\eta_p^2 = .92$, $\varepsilon = .78$, as was the main effect of SOA, $F(1,47) = 1224.45$, $p < .001$, $\eta_p^2 = .96$. The interaction of sentence type and SOA was also significant, $F(2,94) = 259.45$, $p < .001$, $\eta_p^2 = .85$.

Discussion

Experiment 2 investigated whether the immediate processing induced by presupposition triggers is automatic or requires limited capacity. To this end, the PRP approach and the logic of slack-logic were combined with the self-paced reading task already used in Experiment 1.

The ratings largely replicated the results obtained in Experiment 1 with the exception that the difference between neutral and trigger sentences was not significant for the determiner. The overall higher ratings for trigger sentences compared to unacceptable sentences suggest that participants accommodate the presupposition and interpret the sentences (and presuppositions) as intended.

Regarding the tone discrimination task, the main effect of SOA is somewhat unexpected against the background of the central bottleneck model. However, such observations are not uncommon in PRP research. We will come back to this in the General Discussion, where we also consider the overadditive interaction of sentence type and SOA.

Regarding reading times on the trigger position, we obtained several results of interest. First, with a long SOA, the observed pattern of differences between sentence types was the same as in Experiment 1, and thus a successful replication of these results. Secondly, these differences were the same with a short SOA. This result contradicts the notion that the processes initiated by the trigger are automatic. It is noteworthy though that the differences were even larger with a short than with a long SOA—a pattern that is also not predicted by the underlying central bottleneck model. We will come back to this in the General Discussion.

General Discussion

The purpose of this study was to investigate the nature of the immediate processing initiated upon encountering a presupposition trigger. To that end, we compared sentences including a presupposition trigger with grammatical sentences without a presupposition trigger and unacceptable sentences. Experiment 1 replicated several results obtained by Tiemann et al. (2011), but focused only on two triggers, namely definite determiners and *again*. Experiment 2 combined a self-paced reading task with the PRP approach and the locus of slack-logic to address the main question of the present study: Is presupposition processing an automatic or capacity-limited process?

Main Results and Theoretical Implications

Experiment 1 replicated the immediate effects of presupposition processing on the presupposition trigger, as observed by Tiemann et al. (2011). Furthermore, our first experiment revealed that presenting all words preceding the trigger simultaneously and presenting all relevant stimuli in one test session did not influence the immediate effects on the trigger itself or the participants' interpretation.

Experiment 2 was based on the results of Experiment 1 and assessed whether the observed processing initiated by the presupposition trigger is automatic or capacity-limited. The former leads to the prediction that the trigger effects should only be observed with a long but not with a short SOA. This was clearly not the case, and thus an automatic processing of presupposition triggers appears unlikely. Admittedly, the observed overadditive interaction (i.e., a larger trigger effect [of sentence type] with the short than with the long SOA) is also not predicted by the central bottleneck model.[5] However, we can tentatively offer an explanation for this result. In particular, a very similar pattern was observed for the response times in the tone discrimination task. With a short SOA, differences in this task "propagate" into the reading times of the subsequent task, in this case the reading times at the trigger position. This might have induced the overadditivity we observed. Although this is a post hoc explanation, it seems important to again point out that the observed results are certainly not compatible with automatic trigger processing.

The findings are in line with assuming that for presupposition triggers a context search is started immediately: For the definite determiner, the search for an appropriate referent is started, and for *again* a suitable previous event has to be found in the context. It is important to note that this search for a referent must initially be unsuccessful in the presented context, as the presupposition was never explicitly verified, that is, a clear referent for the definite noun phrase was never given and there was no previous event that *again* could refer to either. As a result, participants possibly anticipated to accommodate the presupposition, which the acceptability ratings suggest they did. Accommodation is a poorly understood process, both from the theoretical and experimental point of view. However, under any theory it is usually assumed to be a process of enriching the context with the information of the presupposition, if that is contextually feasible. To check whether adding the presupposition is plausible requires good knowledge of the contents of the context. It is thus unsurprising that in preparation of accommodation cognitive resources like working memory play a role already on the trigger itself.

From a different perspective, assuming a search process is well in line with previous research suggesting a link between presupposition processing and working memory. For example, Anderson and Holcomb (2005) compared test sentences that either included a definite or an indefinite determiner (e.g., "The/A cab came very close to hitting a car.") with a context sentence preceding these test sentences that either introduced the critical noun directly or used a synonym (e.g., "Kathy sat nervously in her cab/taxi to the airport."). The data revealed an enhanced left anterior negativity (LAN) for the

[5] The overadditive interaction remained significant when the neutral sentences were excluded, and thus was not driven by these sentences only.

definite compared with the indefinite determiner, reflecting a referential assignment of the noun phrase following the definite determiner to an antecedent. King and Kutas (1995) also interpreted the increase in the LAN as a referential assignment that increases the demands on working memory (see also Domaneschi et al. 2014). Conceivably, this requires encoding of information into working memory, and exactly this has been shown capacity-limited (Jolicoeur and Dell'Acqua 1998). Furthermore, this assignment is only possible when the relevant entity is found in memory. The necessary search process in turn likely involves repeated selection and de-selection of working memory items that are considered as referents, and selecting working memory items is also a capacity limited process (Janczyk 2017).

Regarding potential similarities and differences in processing between both triggers, the results suggest that both triggers share the feature of capacity-limitations. As suggested, we suspect that the underlying process requiring cognitive capacity is indeed the same. At the same time, however, the numerical difference of the size of the interaction between sentence type and SOA may point to differences. These could result from different difficulties in the underlying search for a potential referent or from an additional process running only for one of the triggers. The present data do not allow for drawing definite conclusions on this matter, however. Moreover, based on our data, a direct comparison between the two triggers is not feasible as we did not control for several factors influencing differences in processing, for example, syntactic position. The present paper offers an indirect comparison by showing that processing both presupposition triggers is capacity-limited. They may require different amounts of capacity and/or different processes; however, the qualitative conclusion is the same for both investigated triggers.

We would like to make two additional comments regarding a comparison. First, a direct comparison between triggers, and especially the one between the definite determiner and *again,* is hardly ever possible, even if many factors are controlled for. This is because both words belong to different categories (determiner vs. adverb), and, as a result, necessarily appear in different syntactic positions and fulfill different syntactic roles. They also occur with different frequency. Second, we believe we did control for several factors that were relevant for our critical Experiment 2. In particular, we tested for naturalness, readability, and predictability by including the acceptability rating task, which should reveal any deviances in these respects. The findings reveal consistent behavior of participants in judging the sentence with the trigger to be acceptable. Further, the role of position was reduced in our experiment as we presented the part before the trigger position simultaneously for all sentence types. In other words, the critical word of investigation always occurred in the second position. Finally, the total word length of the sentence could not have influenced interpretation of the word at the trigger position and, in the critical Experiment 2, this is the only position of interest for the locus of slack-logic.

Limitations and Future Extensions

The probably clearest limitation of the present study is its focus on only two particular triggers in only one language (German). One reason for this choice was to increase the amount of test sentences and thereby improve the precision of aggregate estimates compared to the original study by Tiemann et al. (2011). We purposefully chose triggers belonging to different classes, that is, *again* as a representative of Class 1 and definite determiners as a Class 2 member according to the classification by Tiemann et al. (2015).[6] However, a generalization on the basis of only these two triggers is impossible, and consequently, additional data are needed to gain further insights into potential classes of triggers. Furthermore, future studies should be carried out in different languages than German, to explore the generality of our results across different languages. The important contribution of the present paper is that the methodology and logic of our experiment can be fruitfully extended to investigate difference between (other) triggers and non-triggers, as well as different languages.

Another objection may relate to the choice of the central bottleneck model from which we derived the predictions for Experiment 2. It is certainly true that the adequacy of this model is debated in cognitive psychology. Rather, models have been suggested that allow for parallel processing of the central stage as well, even though the assumption of a capacity-limitation is still made (Navon and Miller 2002; Tombu and Jolicoeur 2003). While parallel processing is possible, the available capacity must be divided between tasks leading to performance decrements in turn. The exact proportion is, for example, determined by a sharing parameter in the model of Tombu and Jolicoeur: if all capacity is first devoted to Task 1, their model mimics the central bottleneck model. The important point here is, however, that the critical predictions for Task 2 (in our case: reading the word at the trigger position) are the same for capacity sharing models and the central bottleneck model. It is simpler, though, to derive and illustrate the predictions from the latter model. Regarding Task 1 performance (in our case: the tone discrimination task), both types of models indeed differ in their predictions. However, the effect of SOA and the overadditive interaction of sentence type and SOA in response times in the auditory discrimination task are more compatible with the capacity sharing approach.

Conclusion

The present study replicates and extends previous results regarding immediate processing of presuppositions, starting on the respective triggers. Two main conclusions can be drawn based on the two experiments we reported: First, encountering a presupposition trigger indeed appears to induce immediate processing of the presupposition. Second, this processing requires limited cognitive capacities and is not automatic.

Acknowledgements Open Access funding provided by Projekt DEAL. This research was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft; DFG) via Project B2 of the Collaborative Research Center 833 "The constitution of meaning" and Project ObTrEx of the DFG Priority Program 1727 "Xprag.de".

[6] There exist other classifications of triggers with which our results are also in line with, for example, informative versus non-informative (Tonhauser 2015), soft versus hard (Abusch 2010), and entailed versus non-entailed (Klinedinst 2016; Sudo 2012).

Appendix

The items used in Experiment 2 were selected on the basis of the ratings obtained in Experiment 1 with regard to two criterions. Criterion 1 was that the difference between the ratings of neutral sentences and trigger sentences was minimal (and at best: negative), and Criterion 2 was that the difference between trigger sentences and unacceptable sentences

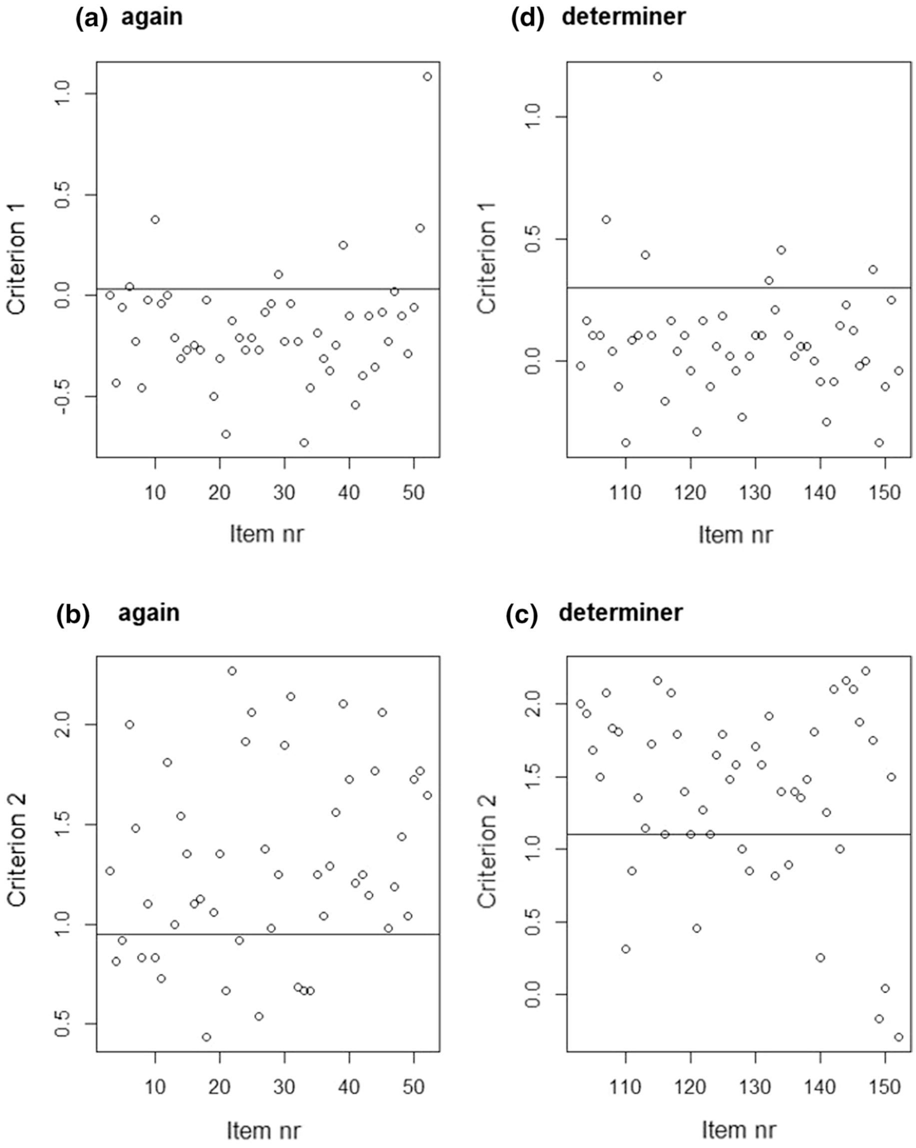


Fig. 8 Visualization of the differences between the relevant sentence types for *again* (**a**, **b**) and determiner (**c**, **d**) that were used to select the 32 items for Experiment 2. The exclusion criteria are visualized with the black line

was maximal. The exact exclusion criteria for the two trigger types were set in a way that approximately one third of the trials were excluded for Criterion 1, and two thirds were excluded due to Criterion 2. For the trigger *again*, exclusion criteria were a value > 0.03 (Criterion 1) or a value < 0.095 (Criterion 2). For the trigger determiner, exclusion criteria were a value > 0.25 (Criterion 1) or a value < 1.00 (Criterion 2) (see Fig. 8 for a visual illustration). The six items for each trigger that were closest to the exclusion criteria were used for the practice block.

References

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes, 24,* 713–734.

Abrusán, M. (2011). Predicting the presuppositions of soft triggers. *Linguistics and Philosophy, 34,* 491–535.

Abusch, D. (2010). Presupposition triggering from alternatives. *Journal of Semantics, 27,* 37–80.

Anderson, J. E., & Holcomb, P. J. (2005). An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain and Language, 94,* 200–216.

Beaver, D., & Geurts, B. (2012). Presupposition. In K. von Heusinger, C. Maienborn & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (pp. 2432–2459). Berlin: Mouton de Gruyter.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language, 98,* 159–168.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *NeuroReport, 18,* 1851–1854.

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review, 14*, 717–722.

Domaneschi, F., Canal, P., Masia, V., Lombardi Vallauri, E., & Bambini, V. (2018). N400 and P600 modulation in presupposition accommodation: The effect of different trigger types. *Journal of Neurolinguistics, 45,* 13–35.

Domaneschi, F., Carrea, E., Penco, C., & Greco, A. (2014). The cognitive load of presupposition triggers: Mandatory and optional repairs in presupposition failure. *Language, Cognition and Neuroscience, 29,* 136–146.

Domaneschi, F., & DiPaola, S. (2018). The processing costs of presupposition accommodation. *Journal of Psycholinguistic Research, 47,* 483–503.

Frege, G. (1892). On sense and reference. In P. Geach & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege* (pp. 56–78). Oxford: Blackwell.

Glanzberg, M. (2005). Presuppositions, truth values and expressing propositions. In G. Preyer & G. Peter (Eds.), *Contextualism in philosophy: Knowledge, meaning, and truth* (pp. 349–396). Oxford: University Press.

Heim, I. (1982). The Semantics of Definite and Indefinite Noun Phrases. Ph. D. thesis, University of Massachusetts.

Heim, I. (1990). Presupposition projection. In R. van der Sandt (Ed.), *Workshop reader "presupposition, lexical meaning and discourse processes: Workshop reader"*. Nijmegen: University of Nijmegen.

Heim, I. (1991). Artikel und Definitheit. In A. V. Stechow & D. Wunderlich (Eds.), *Semantik. Ein internationales Handbuch der zeitgenössischen Forschung. [Semantics. An international handbook of contemporary research]* (pp. 487–535). Berlin: De Gruyter Mouton.

Heim, I. (1992). Presupposition projection and the semantics of attitude verbs. *Journal of Semantics, 9,* 183–221.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. New York, NY: Wiley.

Janczyk, M. (2017). A common capacity limitation for response and item selection in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1690–1698.

Janczyk, M., & Kunde, W. (under review). Dual-tasking from a goal perspective: A review.

Jolicoeur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36,* 138–202.

Jouravlev, O., Stearns, L., Bergen, L., Eddy, M., Gibson, E., & Fedorenko, E. (2016). Processing temporal presuppositions: An event-related potential study. *Language, Cognition and Neuroscience, 31,* 1245–1256.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7,* 376–395.

Kirsten, M., Tiemann, S., Seibold, V. C., Hertrich, I., Beck, S., & Rolke, B. (2014). When the polar bear encounters many polar bears: Event-related potential context effects evoked by uniqueness failure. *Language, Cognition and Neuroscience, 29,* 1147–1162.

Klinedinst, N. (2016). Two types of semantic presuppositions. In K. Allan, A. Capone, & I. Kecskes (Eds.), *Pragmemes and theories of language use. Perspectives in pragmatics, philosophy & psychology* (Vol. 9, pp. 601–624). Cham: Springer.

Kripke, S. A. (2009). Presupposition and anaphora: Remarks on the formulation of the projection problem. *Linguistic Inquiry, 40,* 367–386.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: University Press.

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology, 44,* 193–251.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116,* 220–244.

Piai, V., Roelofs, A., & Schriefers, H. (2014). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 147–165.

Russell, B. (1905). On denoting. *Mind, 14*, 479–493.

Schnur, T. T., & Martin, R. (2012). Semantic picture–word interference is a post-perceptual effect. *Psychonomic Bulletin & Review, 19,* 301–308.

Schwarz, F. (2007). Processing presupposed content. *Journal of Semantics, 24,* 373–416.

Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a Stroop task. *Journal of Mathematical Psychology, 18*, 105–139.

Simons, M. (2001). On the conversational basis of some presuppositions. *Proceedings of Semantics and Linguistics Theory, 11,* 431–448.

Simons, M., Tonhauser, J., Beaver, D., & Roberts, C. (2011). What projects and why. In D. Lutz & N. Li (Eds.), *Proceedings of Semantics and Linguistic Theory* (Vol. 20, pp. 309–327).

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic, 2,* 447–457.

Stalnaker, R. (2002). Common Ground. *Linguistics and Philosophy, 25*, 701–721.

Sudo, Y. (2012). *On the semantics of phi features on pronouns*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology, 14,* 1–36.

Tiemann, S., Kirsten, M., Beck, S., Hertrich, I., & Rolke, B. (2015). Presupposition processing and accommodation: An experiment on wieder ('again') and consequences for other triggers. *Studies in Theoretical Psycholinguistics, 45,* 39–65.

Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., et al. (2011). Psycholinguistic evidence for presuppositions: On-line and off-line data. *Proceedings of Sinn & Bedeutung, 15,* 581–595.

Tombu, M., & Jolicoeur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 3–18.

Tonhauser, J. (2015). Are 'informative presuppositions' presuppositions? *Language and Linguistics Compass, 9,* 77–101.

Tonhauser, J., Beaver, D. I., & Degen, J. (2018). How projective is projective content? Gradience in projectivity and at-issueness. *Journal of Semantics, 35,* 495–542.

Welford, A. T. (1952). The 'psychological refractory period' and the timing of high-speed performance—a review and a theory. *British Journal of Psychology. General Section*, *43*, 2–19.

Wirth, R., Pfister, R., Janczyk, M., & Kunde, W. (2015). Through the portal: Effect anticipation in the central bottleneck. *Acta Psychologica, 160,* 141–151.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

10.4 Study 4

Copyright Notice

Official citation:

Schneider, C. Bade, N., Franke, M., & Janczyk, M. (2020). Presuppositions of determiners are immediately used to disambiguate utterance meaning. *Psychological Research,* https://doi.org/10.1007/s00426-020-01302-7.

The final publication is available at https://link.springer.com/

https://link.springer.com/article/10.1007%2Fs00426-020-01302-7

Psychological Research
https://doi.org/10.1007/s00426-020-01302-7

ORIGINAL ARTICLE

Presuppositions of determiners are immediately used to disambiguate utterance meaning: A mouse-tracking study on the German language

Cosima Schneider[1] · Nadine Bade[2] · Michael Franke[3] · Markus Janczyk[4]

Received: 18 July 2019 / Accepted: 31 January 2020

Abstract

The present study investigated how listeners understand and process the definite and the indefinite determiner. While the definite determiner clearly conveys a uniqueness presupposition, the status of the anti-uniqueness inference associated with the indefinite determiner is less clear. In a forced choice production task, we observed that participants make use of the information about number usually associated with the two determiners to convey a message. In a subsequent mouse-tracking task, participants had to select one of two potential referents presented on screen according to an auditorily presented stimulus sentence. The data revealed that participants use the information about uniqueness or anti-uniqueness encoded in determiners to disambiguate sentence meaning as early as possible, but only when they are exclusively faced with felicitous uses of determiners.

Introduction

Imagine a situation where someone says "The fridge at work is broken". There are probably two assumptions you make when hearing this utterance. First, that there actually exists a fridge at the workplace of the speaker, and second, that there is exactly one fridge. Both of these assumptions are called presuppositions in the semantics/pragmatics literature and they are referred to as the presuppositions of existence and uniqueness of the definite determiner "the".

Presuppositions are background assumptions that speakers of a conversation take to hold. They are usually triggered by a lexical item, the so-called presupposition trigger. In the example above, this would be the definite determiner "the". Since we know that it can only be felicitously used with unique objects due to its presuppositions, we can draw conclusions about the number of fridges based on its use. Compare the example with the utterance "A fridge at work is broken.", that is, with the same utterance containing the indefinite determiner "a". The assumption about the existence of a fridge remains the same. However, there is no inference of the fridge being unique. Rather, one could deduce that there is actually more than one fridge at the workplace of the speaker. Similarly to the case of the definite determiner, we deduce this based on the fact that indefinite determiners are only felicitously used with non-unique objects. However, the status of this *anti-uniqueness* inference is more controversially discussed in the literature.[1] It

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00426-020-01302-7) contains supplementary material, which is available to authorized users.

✉ Cosima Schneider
cosima.schneider@uni-tuebingen.de

Nadine Bade
nadine.bade@ens.fr

Michael Franke
michael.franke@uni-osnabrueck.de

Markus Janczyk
janczyk@uni-bremen.de

1 University of Tübingen, Schleichstrasse 4, 72076 Tübingen, Germany

2 Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, PSL University, CNRS, 29 rue d'Ulm, Paris, France

3 Osnabrück University, Wachsbleiche 27, 49090 Osnabrück, Germany

4 University of Bremen, Hochschulring 18 (Cognium), 28359 Bremen, Germany

[1] Some authors use the term "non-uniqueness" instead of "anti-uniqueness" to refer to this effect of the indefinite determiner (Heim 1991; Sauerland, 2008; but see, e.g., Alonso-Ovalle, Menndez-Benito, & Schwarz, 2011, using the term "anti-uniqueness"). We use the term "anti-uniqueness" to describe the same phenomenon as discussed by these authors, and do not mean to imply any theoretical distinction. Importantly, we do not assume Barker's more specialized definition of "anti-uniqueness" as definite determiners being blocked in partitive constructions (e.g., one of John's friends; Barker, 1998).

Published online: 04 April 2020

has been argued that both uniqueness and anti-uniqueness inferences are part of the non-literal meaning component of sentences. However, open questions are whether they are equal in strength, and how and when these different inferences arise.

The question we address in this paper is whether listeners use inferences associated with definite and indefinite determiners to disambiguate utterance meaning as early as possible, even if the speaker does not always use these determiners felicitously. Much previous research has provided evidence that listeners may, often swiftly, adapt to the idiosyncrasy of the given speaker. This applies to many different aspects of the interpretation of language, including phonetics/phonology (Kleinschmidt & Jaeger, 2015; Roettger & Franke, 2019), syntax (Fine, Jaeger, Farmer & Qian, 2013; Jaeger & Snider, 2013), semantics (Yildirim, Degen, Tanenhaus, & Jaeger, 2016), and pragmatic factors (Grodner & Sedivy, 2011; Stranahan, 2018). However, to the best of our knowledge, there is so far no work investigating adaptation effects of presuppositional information for online processing. Moreover, there is a vast amount of literature on the processing of felicitous and infelicitous uses of presuppositions in different contexts (see Schwarz 2007; Schwarz & Tiemann, 2012; Tiemann et al., 2011; Tiemann, 2014; for a recent review, see Schwarz, 2016), but almost no experimental investigations of when and how the interpretation of utterances may be affected by information encoded in the presuppositions of certain expressions. The present study tries to fill these gaps. To this end, we will compare how listeners' online interpretation of definite and indefinite determiners changes depending on how reliable the speaker is in using the determiner felicitously. To do so, we compare a group of participants encountering only felicitous uses of determiners with a group that also encounters infelicitous uses of determiners. In the following, we will introduce presuppositions and the triggering mechanism in more detail, focusing on determiners. This is followed by a discussion of earlier work suggesting that the process of presupposition evaluation is started immediately on the presupposition trigger. We report one experiment comprising two parts: in a production task, we test whether participants use the presupposition usually associated with the determiner to convey information. The subsequent mouse-tracking task addresses whether participants use determiners to disambiguate utterance meaning and if so, whether this is different for the definite and the indefinite determiner.

Presuppositions

Informally speaking, presuppositions are background assumptions that are shared by all interlocutors of a conversation. They are introduced by certain words, so-called presupposition triggers. One classical example of such a trigger is the definite determiner. It introduces the presupposition that there exists a unique individual with the property described by the noun it combines with. Technically speaking, these words introduce appropriateness conditions, that is, certain restrictions on what the context must look like for the sentence containing them to be felicitously uttered. The assumption is that if the presupposition of a trigger is not met in the context, uttering a sentence containing it is infelicitous. This is illustrated in (1). In (1-a), the presupposition of the definite determiner, that there is a unique apple, is fulfilled in the context. Thus, the sentence is felicitous. In (1-b), the presupposition of uniqueness is not fulfilled in the context and the sentence is infelicitous (or inappropriate).

1. (a) Context: There is an apple and a banana on the desk.
   Please give me the apple.
   (b) Context: There are three apples and a banana on the desk.
   # Please give me the apple.

Whereas felicitous uses of the definite determiner require a unique discourse referent (i.e., the definite determiner presupposes uniqueness), the indefinite determiner is assumed to require there to be more than one referent in the context. As a result, its use becomes odd if it is known that the referent is unique, as in (2).

2. # A sun is shining.

It was thus suggested in the literature that the indefinite determiner presupposes anti-uniqueness (see Kratzer, 2005, or the discussion in Heim, 1991, 2011) in the same way the definite determiner presupposes uniqueness. Under this view, both definite and indefinite determiners come with their own restrictions on what are appropriate contexts. Henceforth, we will refer to this theory as the "presupposition theory" and assume that both inferences are equally robust and accessed quickly.

However, Heim (1991) noted that (3) can be uttered without it being certain that there is more than one 20-ft-long catfish: it suffices that the speaker is not sure that there is exactly one 20-ft-long catfish.

3. Robert caught a 20-ft-long catfish.

From such observations, Heim (1991) concluded that the inference associated with the indefinite determiner is weaker than the presupposition of the definite determiner. To capture this, she proposed to add another principle to the Gricean maxims of conversation, Maximize Presupposition, which says: Presuppose as much as possible! (see also Chemla, 2009; Percus, 2006; Sauerland, 2008; Schlenker, 2012, for more refined versions of Maximize Presupposition). This principle can account for the fact that using indefinite determiners is infelicitous when it is common ground that the referent is unique, as in (4-a). More specifically, it explains the oddness of (4-a) by means of pragmatic reasoning over presuppositional stronger alternatives (Heim, 1991; Percus, 2006; Sauerland, 2008; Schlenker, 2012). The sentence in (4-b) is an alternative since it only differs from its competitor regarding the presuppositions it introduces. It introduces more presuppositions that are true in the context (since we know people have one unique father) and is thus the presuppositionally stronger alternative. When a listener hears the presuppositionally weaker sentence (4-a), s/he assumes the speaker must believe the presupposition of the stronger alternative to be false. The reasoning behind this conclusion is based on two main assumptions: (1) that the speaker obeys the conversational maxims including Maximize Presupposition, and (2) that the speaker tries to be cooperative in doing so. The hearer thus assumes that if the speaker believed the presupposition of (4-b) to be true, s/he would have used this version, because it would be more informative on a presuppositional level. Since s/he did not, s/he must believe it to not hold. The belief that the victim does not have a unique father, however, is contrary to common knowledge and, therefore, leads to the oddness of (4-a).[2]

4. (a) ∗ A father of the victim arrived at the crime scene. (Heim, 1991)
   (b) The father of the victim arrived at the crime scene.

In other words, anti-uniqueness is derived by (1) considering the (stronger) alternative with the definite determiner and (2) negating its presupposition. The inferences, which are the result of pragmatic reasoning based on Maximize Presupposition, are not presuppositions proper under this view. Henceforth, we will refer to them as anti-presuppositions (Percus, 2006). They are theoretically kept apart from presuppositions (and implicatures) and should be processed differently. As opposed to the assumptions of the "presupposition theory" introduced above, the anti-presupposition arising with the indefinite determiner has a weaker status than the presupposition of the (stronger) definite determiner; we will refer to this theory as the "anti-presupposition theory" in the following. Because the anti-uniqueness inference is derived by initially considering the (stronger) alternative of the definite determiner and subsequently negating it, processing of the indefinite determiner should thus be more complex than processing the definite determiner.

A third type of theory assumes that the indefinite determiner triggers an implicature due to its competition with other quantificational terms, for example, "every/all" or "another" (Chierchia, Fox, & Spector, 2012; Grønn & Sæbø, 2012). These quantificational terms form a lexical scale with the indefinite determiner (Horn, 1972). An implicature arises when the weaker item on such a scale is chosen, in this case the indefinite determiner. All items that are higher in the scale (items that trigger stronger alternatives) get negated: for example, the implicature of "A boy came" is that "Not all boys came". This negation process requires the assumption of existence of other boys, and anti-uniqueness follows as a consequence. Contrary to the competition on a presuppositional level according to Maximize Presupposition, the competition between alternatives in this case arises on the level of assertion ("a" and "all" differ on the level of assertion, whereas "a" and "the" are alike on that level). Variants of this argue that anti-presuppositions *are* essentially implicatures in that they can be informative and follow from the same general mechanism (of exhaustification) (Magri, 2009; Schlenker, 2012; Singh, 2011). According to these theories, the definite and indefinite determiners are also asymmetric in the inferences they introduce: whereas the indefinite determiner should come with an implicature, which has shown to be processed even more rapidly (at least if certain conditions are met) than presuppositions (Bill, Romoli, & Schwarz, 2018; but see also Chemla, 2008), the definite determiner should come with a presupposition. We will refer to this theory as the "implicature theory".

In sum, three different approaches are available to explain the effects of uniqueness and anti-uniqueness resulting from the definite and the indefinite determiner. (1) According to the "presupposition theory", both determiners carry their own presuppositions proper and thus processing the two determiners should be equally difficult. (2) According to the "anti-presupposition theory" based on Maximize Presupposition, the anti-uniqueness inference of the indefinite determiner is derived indirectly from negating the uniqueness presupposition of the definite determiner. Thus, processing indefinite determiners should be more difficult than processing a definite determiner. (3) According to the "implicature theory", the indefinite determiner comes with an implicature instead of an (anti-)presupposition. In this case, one may expect the indefinite determiner to be more easily processed than the definite determiner.

Previous investigations of processing determiners

Experimental investigations of presuppositions have increased in recent years (for a recent review, see Schwarz,

[2] We are aware that the example is outdated, as nowadays gay couples frequently adopt children, which seems to lead to language change, especially in lexical fields dealing with family concepts. We think that the example still illustrates the relevant point well though, if we imagine the noun phrase to refer to the biological father, not the person(s) fulfilling the role of a father.

2016). Here, we focus on studies dealing with (1) the early processing of presuppositions/inferences triggered by determiners, (2) differences between felicitous and infelicitous uses of determiners, and (3) potential differences between definite and indefinite determiner.

In a self-paced reading study, Altmann and Steedman (1988) investigated the syntactic consequences of a definite noun phrase having its presuppositions met or not met by the context. The data revealed early processing of the presupposition, before the end of the sentence. More precisely, participants were presented with test sentences in two different contexts. Context 1 introduced two candidates for a potential referent (a safe with a new lock and a safe with an old lock, see (5)), while Context 2 introduced exactly one candidate for a potential referent (see (6)).

5. Context 1: A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was a safe with a new lock and a safe with an old lock.
6. Context 2: A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was a safe with a new lock and a strongbox with an old lock.

In the test sentence in (7), the prepositional phrase "with the new lock" modifies the noun phrase "safe". As a result, the uniqueness presupposition is met in both contexts for this test sentence. This is not the case for the test sentence in (8), whose presupposition is only satisfied by Context 2.

7. The burglar/blew open/the safe/with the new lock/and made of/with the loot.
8. The burglar/blew open/the safe/with the dynamite/and made of/with the loot.

Reading times differed in the disambiguation region (i.e., on the prepositional phrase "with the new lock" or "with the dynamite"). Test sentences with an unmet uniqueness presupposition as in (8) were read slower than test sentences as in (7). Thus, the authors conclude that people experience processing difficulties at an early point in time, when the uniqueness presupposition of the definite determiner is not met. However, no evidence for processing difficulties on the presupposition trigger itself was reported. We believe that the relatively late effects are most likely due to the experimental design. In particular, the content of the presuppositions was only known on the prepositional phrase, because only then it was clear which referent was considered unique in the context. In sum, the results suggest that a presupposition is processed as soon as it is fully known. A more detailed analysis of syntactic ambiguity resolution strategies was done by Spivey, Grosjean, and Knoblich (1995) who provide further evidence for an immediate influence of pragmatics and logically specific biases in syntactic ambiguity resolution.

Tiemann et al. (2011) reported three self-paced reading studies and acceptability ratings on presupposition processing as induced by different triggers (German "wieder": English "again"; "auch": "also"; "aufhören": "stop"; "wissen": "know", and definites in the shape of possessive noun phrases). In their first experiment, which is important for the current paper, they focus on the processing of the trigger itself. Participants were presented with a context (as in (9)) and the authors compared test sentences including a presupposition trigger (as in (10)) with sentences including a neutral word that does not trigger a presupposition (as in (11)), and with semantically unacceptable sentences (as in (12)).

9. Context: Tina ist mit einer guten Freundin shoppen.
   Tina is shopping with a good friend.
10. Sie kauft wieder rote Handschuhe.
    She buys red gloves again.
11. Sie kauft heute rote Handschuhe.
    She buys red gloves today.
12. *Sie kauft freundlich rote Handschuhe.
    She buys red gloves friendly.

Sentences with the neutral word were rated best, followed by sentences including the presupposition trigger and unacceptable sentences. Second, and most importantly, reading time data revealed that—for the position of the presupposition trigger—sentences with a trigger induced the longest reading times, followed by sentences with a neutral word, while unacceptable sentences were read fastest (see also Schneider, Bade, & Janczyk, 2020). This pattern suggests that a presupposition trigger immediately demands more attention, because it alerts the reader to consider the preceding context. Furthermore, in their third experiment, they investigated whether a presuppositional sentence in a neutral context (neither making the presupposition explicitly true nor false) evokes longer reading times than in a falsifying or verifying context. The data revealed early effects on the trigger, which also suggest that processing of the presupposition begins immediately upon encountering the trigger.

Unfortunately, reading times were not analyzed for individual presupposition triggers, as there were not enough items for each trigger to allow for strong conclusions. It is thus unclear whether the same pattern holds for all triggers or not.

There is further evidence for an immediate processing of presupposition triggers from electrophysiological studies. For example, van Berkum, Brown, and Hagoort (1999) and van Berkum, Brown, Hagoort, and Zwitserlood (2003) used event related potentials (ERPs) to investigate the interplay of referential and structural factors during sentence processing in discourse. To do so, the authors used referentially ambiguous noun phrases. In these studies, participants were presented with a discourse as in (13) and (14). A corresponding test sentence was, for example, "David told the girl that had been on the phone to hang up." In discourse (13), uniqueness of the noun phrase "the girl" in the test sentence was granted, because the discourse introduces only one salient girl. In contrast, this is not the case in the other discourse (14), where both girls are equally salient. As a result, the uniqueness presupposition is not fulfilled up to the disambiguating relative clause ("...that had been on the phone...").

13. David had told the boy and the girl to clean up their room before lunchtime. But the boy had stayed in bed all morning and the girl had been on the phone all the time.
14. David had told the two girls to clean up their room before lunchtime. But one of the girls had stayed in bed all morning and the other girl had been on the phone all the time.

The definite noun phrases evoked early ERP effects already on the noun, when the uniqueness presupposition was not met. Thus, referential ambiguity appears to be detected very early during sentence processing. Contrary to Altmann and Steedman (1988), evaluation of the presupposition was not delayed until to the relative clause, but instead participants considered the presupposition against the context as early as they heard or read the noun. This may come as a surprise given that participants knew that there were also sentences where the presupposition failure was resolved by the following relative clause, similar to the participants in Altmann and Steedman (1988) becoming aware that the following prepositional phrase was important for evaluating the presupposition. The difference might be due to the different syntactic status of relative clauses and prepositional phrases. Taken together, the studies of van Berkum et al. (1999, 2003) support the idea of early processing of presuppositions triggered by the definite determiner. Participants realize that there might be presupposition failure of a definite noun phrase immediately on the noun itself.

Evidence for immediate presupposition processing also comes from Kirsten et al. (2014) who investigated the processing of definite and indefinite determiners in an ERP study. Two types of contexts (see (15)) introduced either a single referent (e.g., one polar bear) or multiple referents (e.g., some polar bears). Test sentences were alike except for the determiner used ("the/a") and were either presented in a matching condition where the context sentence introduced the noun phrase with an indefinite determiner "ein/e" (Engl.: "a") or in a mismatching condition where it contained a quantifier such as "einige" (Engl.: "some") or "viele" (Engl.: "many").

15. (a) Antje war gestern im Zoo in Düsseldorf und besuchte einen /einige Eisbären im Bärengehege.
    (b) Antje visited the Düsseldorf zoo yesterday and saw a/some polar bear/s in the bear enclosure.
16. (a) Antje beobachtete, dass der/ein Eisbär sehr aggressiv war.
    (b) Antje noticed that the/a polar bear was very aggressive.

The data revealed that participants recognized mismatching conditions already when reading the determiner. For both determiners, the mismatching effect became visible through an N400 and a P600 effect after onset of the noun.[3] Thus, the results support the idea of immediate processing of presuppositions, already starting on the trigger.

There is further evidence that information encoded in determiners is exploited to guide behavior. Dahan, Swingley, Tanenhaus, and Magnuson (2000) focused on gender information encoded in the determiner (in the French language) in an eye-tracking study and demonstrated that gender-marked determiners immediately directed the listeners' eyes towards the object that matched the gender. This supports the idea of immediate processing of determiners and the use of information encoded in therein.

Further support for early processing of determiners comes from a visual-world eye-tracking study using a picture selection task (Bade & Schwarz, 2019a). In one critical condition, sentences like "A/The shirt in Benjamin's closet is blue" were presented auditorily and paired with three different pictures, all of which showed a boy with a closet. On one of the pictures, the closet contained three shirts, one of which was blue (non-unique condition). On another picture, only one blue shirt was depicted (unique condition), and no shirts were depicted on a third (distractor) picture. Participants were asked to choose the picture they thought was corresponding to the sentence, and indeed they looked at the respective target picture (picture with a single shirt for the definite determiner, and picture with multiple shirts for the indefinite determiner) very early on upon hearing the noun. This suggests that the (anti-)uniqueness information encoded in the determiners was used rapidly for interpretation. In addition, differences between both determiners were observed. Inferences based on the use of the indefinite determiner were drawn to a much lesser degree than those evoked by the use of the definite determiner, as demonstrated by fewer target choices for the indefinite than for the definite determiner. Further differences between the determiners

[3] The authors describe the N400 effect as a detection of a semantic mismatch (see, e.g., Kutas, Petten, & Kluender, 2006) and the P600 as an index for a subsequent reanalysis process (see, e.g., Kuperberg, 2007).

were observed in eye-tracking patterns for the cases where the target was chosen. Overall, these results are in line with the "anti-presupposition theory", which predicts differences in processing patterns associated with the anti-uniqueness inference and the uniqueness presupposition.

Finally, mouse-tracking data from Schneider et al. (2019) also support this view. In two experiments, participants were asked to judge the appropriateness of sentences like "Of these, Jan received the/a banana." in contexts showing that Jan's mother bought three pieces of fruit (e.g., one banana and two pears, or two bananas and one pear). Participants made their judgment by moving the mouse cursor into response boxes located in the top right and left corners of the computer screen. The indefinite determiner was associated with more difficulty in processing (reflected in longer movement times and a larger area under the curve). Most importantly, the data of Schneider et al. (2019) also revealed an initial deviation into the direction of the non-target response (i.e., the competitor) for the indefinite determiner. This was predicted by the "anti-presupposition theory" exploiting Maximize Presupposition (Heim, 1991), which suggests that participants first consider the uniqueness presupposition of the definite alternative when encountering an indefinite determiner.

In sum, there is evidence for early processing of presuppositions introduced by definite determiners, as demonstrated by effects observed on the trigger and briefly thereafter, when the content of the presupposition is known. Furthermore, additional processing costs were observed when the presupposition is not met by the context. This effect may, however, be modulated by additional cognitive load (see Clifton, 2013). Finally, the available evidence suggests processing differences between determiners.

Overall, the existing experimental literature on presuppositions focused on the processing costs associated with presuppositions of different triggers in different contexts. One question that this research concentrated on was whether presuppositions are more difficult to process in contexts in which their use is infelicitous (i.e., when the presupposition was not verified). Another focus was on the question, when these effects occur. With the exception of Bade and Schwarz (2019a, b) previous research did not address the issue of whether early information about uniqueness and anti-uniqueness encoded in determiners is used to make predictions about a sentence's meaning. A weakness of the few studies that addressed how inferences associated with definite or indefinite determiners affect interpretation (Bade & Schwarz, 2019a) is that it was unclear from the task used in the experiment whether the information encoded by the determiner would be relevant for the choice. In general, it was unclear for the participants what was hinging on their choice of picture. Another weakness of many experiments on definiteness in general is that it only became apparent on the noun whether uniqueness or anti-uniqueness was satisfied, because more than one referent to which the definite or indefinite noun phrase could refer was provided in the context.

The experiment reported in the present paper extends the still small empirical basis for answering the following question: Is interpretation driven by presuppositional information encoded in determiners? An advantage of the present study over previous ones is that disambiguation was possible on the determiner itself, and not only on the noun. This makes it possible to identify immediate effects of the number information encoded in the determiner. Another advantage of the study presented here is that participants were directly addressed by the speaker. The assumption was that hearers would use all available cues given that they had to do a task for the speaker. Moreover, participants were informed that the speaker shared the same knowledge. As opposed to previous studies, participants thus had an incentive to draw the relevant inferences, and were informed of the epistemic state of hearer and speaker they needed to assume.

Experiment

In this section, we will first provide a forecast of the experimental approach followed by a brief introduction into mouse-tracking, the method we used to answer our main questions. We will then lay out the hypotheses in more detail.

The entire experiment comprises two parts. The first part is a forced choice production task where participants are asked to produce a sentence which appropriately describes a given situation. This task has several purposes: First, we aim at showing that participants are aware of the uniqueness presupposition of the definite determiner, and therefore use it predominantly in situations where the described item is unique. Second, we aim at testing whether participants systematically use the indefinite determiner when the object they want to refer to is non-unique. Finally, the data from the production task will be used to screen participants regarding whether they have a sufficient command of the correct usage conditions of both determiners in sentence production. We will exclude participants who commit more than 20% errors, that is, who do not use determiners in the intended way or make errors with the color or object choice (these participants will be replaced with new participants).

The second part is a mouse-tracking experiment with the aim to test whether listeners can rapidly integrate potential cues about uniqueness or anti-uniqueness in a context to achieve early predictive disambiguation, even before hearing the lexically disambiguating referent noun.

Mouse-tracking

Mouse-tracking has become a common method in cognitive psychology, since recent studies revealed that motion trajectories can reflect underlying cognitive processes. In fact, simple hand movements offer a continuous stream of motor output and provide real-time read-outs of ongoing cognitive processes. Spivey et al. (2005) were among the first who used mouse-tracking to answer language-related questions. Participants were instructed to start a trial via moving the mouse into a start box (in the lower center of the screen) and then follow the instructions of an auditory stimulus sentence (e.g., "Click on the candle!") while watching a picture with items in the upper left and upper right corners of the screen. In one condition, the depicted words were similar in their initial phonemes (e.g., "candle" vs. "candy"), while in the other condition they were not (e.g., "candle" vs. "summer"). The trajectories of the movements towards the correct upper corner showed an attraction of the distractor word, when both words shared the initial phonemes. Thus, during processing of the target word, competing phonological representations appear active and influence the exact way the hand moves. Similar approaches have since been applied to, for example, social cognitive questions (Freeman, Dale, & Farmer, 2011), conflict tasks (Scherbaum, Dshemuchadse, Fischer, & Goschke, 2010), the effects of irrelevant stimulus variation on action execution (Janczyk, Pfister, & Kunde, 2013), or the influence of actions consequences on action execution (Pfister, Janczyk, Wirth, Dignath, & Kunde, 2014). Furthermore, mouse-tracking has also been used in sentence verification tasks to study conversational implicatures (Sauerland, Tamura, Koizumi, & Tomlinson, 2015; Tomlinson, Bailey, & Bott, 2013), predictive disambiguation based on early intonational cues (Roettger & Stöber, 2017; Roettger & Franke, 2019), and sentence negation (Dale & Duran, 2011).

Several parameters are usually extracted from the trajectories for further statistical testing. We here focus on the following parameters (see Fig. 1 for an illustration): (1) area under the curve (AUC) is the geometric area between the observed mouse-trajectory and an idealized straight line and becomes larger, the more the trajectory deviates from the straight line. According to Freeman and Ambady (2010), AUC provides a general measure of processing difficulty (Farmer, Cargill, & Spivey, 2007; Freeman, Ambady, Rule, & Johnson, 2008; Spivey, 2008). (2) Movement time (MT) is the time from stimulus onset until reaching the target box. (3) Turn towards target (TTT) is defined as the point in time when participants finally make their decision and turn towards the target without any subsequent reversals (Roettger & Franke, 2019).

In the experiment proper, participants were asked to hand a named object from a shelf to the speaker. One or two

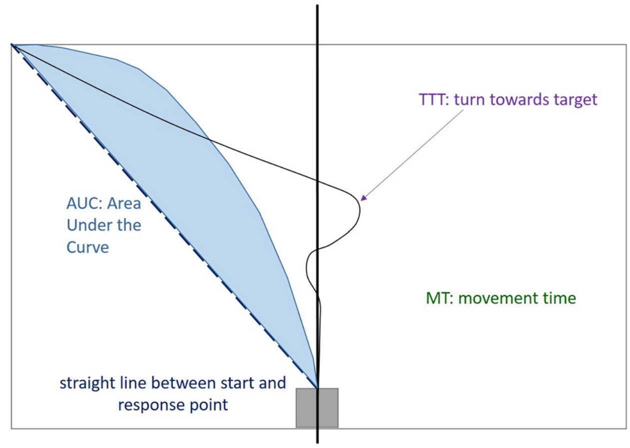


Fig. 1 Illustration of parameters extracted from the mouse-trajectories used for further statistical analyses

items of the named object type were present on the shelf. In addition, one or two entities of a different object type (the competitor) were present. Stimulus sentences either contained the definite or the indefinite determiner. In "early disambiguation" conditions, participants could, in principle, know which object will be referred to already upon hearing the determiner. In contrast, in "late disambiguation" conditions, this is only possible after the target object has been named. Conceivably, however, early disambiguation makes sense only if a listener knows by experience that sentences with definite and indefinite determiners are used in a felicitous way. Thus, one group of participants only encountered felicitous sentences (the reliable group), a second group was presented with infelicitous sentences as well (the unreliable group).

An advantage of using German stimuli is that we can make sure that the definite description could not be plural, as the definite determiner is marked for grammatical gender. As the plural determiner for all three genders is form-identical with the singular feminine determiner, we only used masculine and neuter nouns in our experiment, which makes the definite determiner unambiguously singular. This is not the case in English, where "the" could still be followed by plural nouns, which would make the sentence completely acceptable in non-unique scenarios. The presupposition of definite determiners is that there is a unique maximal element, which is trivially fulfilled for the extension of any plural noun (as there always must be a unique maximum) (cf. Heim, 2011). Only in combination with the singular does the definite noun phrase presuppose uniqueness of an atomic individual with the described property. It is thus possible to predict on the (singular marked) definite determiner that the reference will fail if there are only non-unique referents in German, but not in English, where plural/singular marking only becomes obvious on the noun. The general expectation is thus that participants in our experiment could use the information of number encoded in the determiner right away and make the according choices rapidly.

Hypotheses

The first question we ask is whether people utilize the relevant meaning components of a determiner (uniqueness and anti-uniqueness) to form expectations about the likely referent, even before this referent is lexically given (and thus disambiguated). This is supported by previous experiments showing that presupposition processing starts on the trigger itself (e.g., Tiemann et al., 2011; van Berkum et al., 1999, 2003). Thus, for the "reliable group", we hypothesize that conditions with unequal amounts of different objects on the shelf (i.e., the early disambiguation conditions) allow faster decisions already upon encountering the determiner in comparison with those conditions with equal amounts of different objects on the shelf (i.e., the late disambiguation conditions; Hypothesis 1).

A second question is whether there are processing differences between the definite and the indefinite determiner. The three theories introduced in Sect. "Presuppositions" allow three different predictions: (1) according to the "presupposition theory", both determiners come with their own presuppositions. In this case, the prediction would be that both the uniqueness and the anti-uniqueness presupposition are accessed equally fast and cause comparable processing difficulties. Thus, no differences are expected in this case (Hypothesis 2a). (2) According to the "anti-presupposition theory", processing the indefinite determiner requires an initial consideration of the uniqueness presupposition of the definite determiner and its subsequent negation. This predicts more processing difficulties for the indefinite than for the definite determiner (Hypothesis 2b; see also Schneider et al., 2019). (3) Finally, the "implicature theory" assumes that the indefinite determiner activates a different type of competition, namely on the level of assertion. This would make the associated inference an implicature, which is processed more rapidly than a presupposition. Accordingly, less processing difficulties are predicted for the indefinite than for the definite determiner (Hypothesis 2c; see also Bill et al., 2018).

A final question is whether the potential early effects of information about uniqueness and anti-uniqueness (as suggested in Hypothesis 1) are affected by occasionally infelicitous uses of the determiner.

The theories spelled out above make different predictions regarding how processing of either determiner is influenced by occasional infelicitous uses. According to the anti-presupposition theory, the definite determiner comes with a lexically stored presupposition whereas the indefinite determiner's inference is the result of pragmatic reasoning. As a consequence, the definite determiner should be less affected by infelicitous uses than the indefinite determiner given the assumption that lexical information is generally harder to overwrite. Following the presupposition theory, both determiners come with lexically encoded information regarding number. Therefore, they should be equally affected by infelicitous uses. Finally, in the case of the implicature theory, the indefinite determiner should be more affected than the definite determiner. This is because implicatures are highly context-dependent expressions, easily affected by speaker reliability (Bott & Noveck, 2004), as opposed to lexical presuppositions.[4] These hypotheses will be referred to as Hypothesis 3.

Method

Participants

The intended sample size was $n = 60$. In total, data were collected from 76 people from the Tübingen (Germany) area. We excluded 14 participants that committed more than 20% errors in the production task, one participant was not a native speaker of German, and one additional participant was excluded because of technical problems during data recording (final sample: mean age = 23.6 years, 48 females, 12 males). All participants in the final sample were native speakers of German, reported normal or corrected-to-normal vision and hearing abilities, and signed written informed consent prior to data collection. Participants were randomly assigned to either the reliable group or the unreliable group of the mouse-tracking part of the experiment, with $n = 30$ per group.

Apparatus and stimuli

Stimulus presentation and response collection was controlled by a notebook connected to a TFT-screen (resolution: 1280 × 1024 px.; visible screen size: 53 × 30 cm). A standard computer mouse was used with slightly reduced cursor speed and non-linear acceleration turned off.

At the top of each screen, a shelf with ten compartments was visualized. The left- and the rightmost compartment contained either one or two objects (the same within one compartment, but different objects on the left and right side). A total of 19 different objects was used, each object in two different colors.

[4] A preregistration report including the hypotheses, the planned design of the experiment, the stimulus material, and the planned analyses is available at OSF: https://osf.io/aym9p/. Parts of the preregistration were written in past tense, as if the text were a method section proper. We regret that this is somewhat unfortunate but confirm that data were collected only after posting the preregistration.

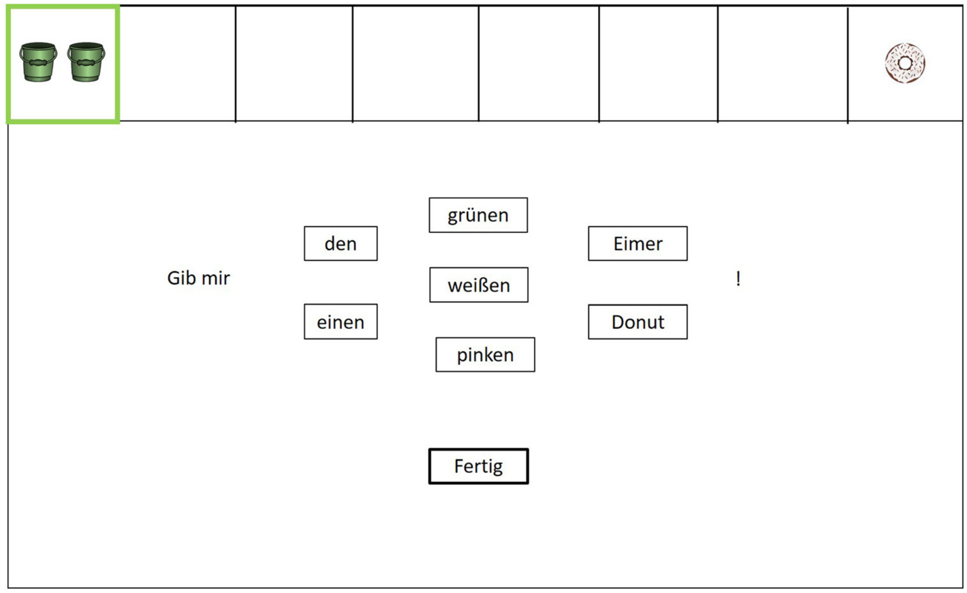


Fig. 2 Example of the stimulus setup for a trial in the production task

During the production task, additional elements were visible on the screen on each trial (see Fig. 2). First, either the left- or the rightmost compartment was outlined in green color to indicate the relevant type of object in the current trial. Second, the German words "Gib mir" (Engl.: "Give me") indicated the start of the to-be-completed sentence. Third, three rows of boxes contained the words the participants were to choose to complete the required sentence (left row: determiner; middle row: adjective (i.e., color of the object); right row: noun (i.e., the object)). In the production task, objects were (white or pink) donuts and (green or white) buckets. An additional box below contained the German word "Fertig" (Engl.: "Done"). Clicking on a box turned the frame bold to indicate selection.

Four different conditions of required sentences were constructed (see Table 1), resulting from combining two variables. (1) The variable *determiner* captures whether the correct determiner would be the definite or the indefinite determiner. (2) The variable *disambiguation* indicates whether an early part of the sentence (i.e., the determiner) can be used to disambiguate the sentence meaning, or only a late part (i.e., the noun). Early disambiguation is possible when different numbers of target and of the competitor are present; only a late disambiguation is possible when the same number of target and competitor is present.

For the mouse-tracking task, an additional box was present centrally at the bottom part of the screen as the start box (see Fig. 3). The left- and rightmost compartments of the shelf served as the response boxes. The remaining 17 objects (i.e., excluding the donut and bucket, see Table 5 in the Appendix) were used in the mouse-tracking task. In the following, we will refer to the object mentioned in the auditory stimulus sentence as the target. The other object, appearing in the opposite compartment of the shelf, will be referred to as the competitor. Some restrictions apply to the construction of the possible trials. First, the common nouns for all competitors had the same grammatical gender as those for the target, to avoid early disambiguation by this information. Second, for the same reason, targets and competitors were always presented in the same color. Third, no nouns with feminine gender were used, because the German feminine determiner "die" could also occur in combination with a plural object (e.g., "die roten Kerzen"; Engl.: "the red candles").

Stimulus sentences as in (17) were pre-recorded and delivered via headphones. These auditory stimuli were constructed in a way to keep prosodic characteristics as constant as possible. To this end, in a first step, all sentences were recorded. Then the sentences were cut and put back together, with the same initial part ("Gib mir den/das"; Engl.: "Give me the") for all stimuli including the definite determiner and the same initial part ("Gib mir ein/einen"; Engl.: "Give me a") for all

Table 1 Summary of all four types of conditions for the production task. In the examples, we assume that the required object is located in the leftmost compartment of the shelf. (Note: determ. = determiner; disamb. = disambiguation; def. = definite; indef. = indefinite)

	determ.	disamb.	target	competitor	sentence
1	def.	early	1 bucket	2 donuts	Give me **the** green bucket
2	def.	late	1 bucket	1 donut	Give me **the** green bucket
3	indef.	early	2 buckets	1 donut	Give me **a** green bucket
4	indef.	late	2 buckets	2 donuts	Give me **a** green bucket

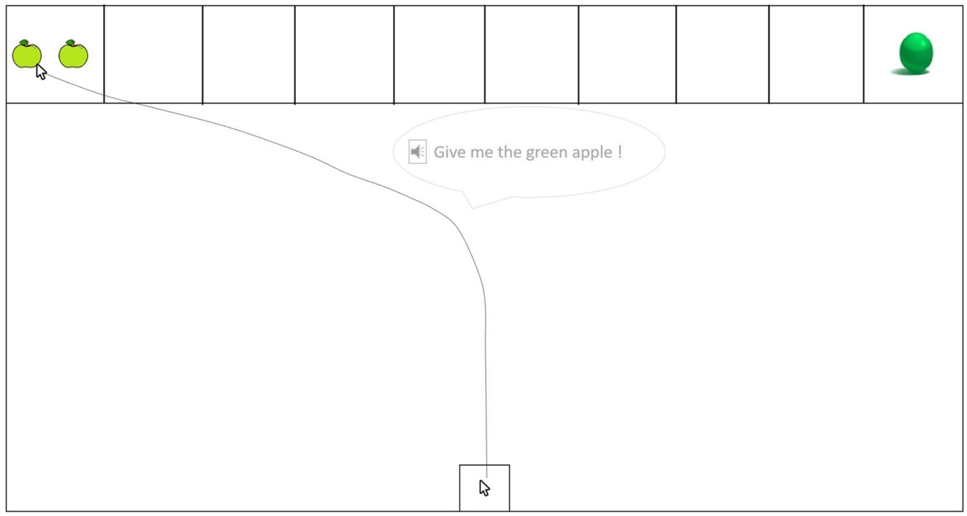


Fig. 3 Example of the stimulus setup for a trial in the mouse-tracking task

stimuli including the indefinite determiner. The remainder of the sentence—for example, "grünen Apfel" (Engl.: "green apple")—was the same for definite and indefinite conditions. It was combined with each of the beginnings to keep phonological differences between sentences with definite and indefinite determiners minimal. For every second item, either the sentence with the definite determiner or the sentence with the indefinite determiner was used as the basis of cutting and pasting. An example of the setup for the mouse-tracking is given in Fig. 3.

17. Gib mir den / einen grünen Apfel!
    Give me the / a green apple!

Each relevant pairing of visual display and sentence can be classified along three variables with two levels, thus resulting in eight possible conditions (see Table 2 for a summary). (1) The variable *determiner* captures whether the sentence comprises a definite or indefinite determiner. (2) Based on the objects in the shelf, the number information carried by the determiner can either be used to disambiguate early or not; this is captured in the variable *disambiguation*: disambiguation could, in principle, happen early (on the determiner) or only late (on the noun). (3) Finally, the variable *felicity* captures whether the use of determiner and head noun was felicitous given the objects in the shelf, that is, whether the number information carried by the determiner about the head noun was actually true.

To create different instances of these relevant experimental conditions, we took the 17 objects, 5 of which were neuter and 12 were male in gender, and each was instantiated in one of two colors. We then created all possible instances of each of the eight experimental conditions by picking one object as target and another as the competitor, such that target and competitor had the same gender and the same color. Pictures instantiating the relevant experimental conditions were then created by showing one or two pictures of the target object together with one or two objects of the competitor object, depending on the requirements of the condition to be instantiated. We then created the sentence belonging to this condition, using either the definite or indefinite determiner, always including the name of the target object, of course. In this way, 138 instances of each condition were created. In the experiment, participants saw random instances of each condition, such that no participant saw the same instance of a condition twice.

Task and procedure

All participants started with the production task and then performed the mouse-tracking task. For the *production task*, participants had to build a sentence to describe the item in the box, which was highlighted by a green frame. In other words, they were to provide an accurate description of the highlighted object. To this end, they were instructed to describe the picture to a second interlocutor, who views the same shelf with objects, but who is not able to see the highlighted box. In addition, participants were told that it is only possible to take one object out of the shelf.

Participants selected words with a left mouse-click within the respective boxes. When a word from all three columns was selected, a click on the "Fertig"-box initiated the next trial. Once made, a selection could not be changed. The two objects, their colors, and the different possible quantities resulted in 32 different trials, which were presented in randomized order. The position of the target (i.e., the left- or rightmost compartment of the shelf) was determined randomly for each trial with the restriction that the target appeared 16 times on each side.

The participants' task in the mouse-tracking task was to select one of the two potential referents presented in the left- and rightmost shelf compartment according to the auditorily presented stimulus sentence. To this end, the participants

Table 2 Summary of all eight possible experimental conditions for the mouse-tracking task

	determ.	disamb.	felicity	sentence	#target	#competitor	picture
1	def.	early	fel.	Give me **the** green apple!	1	2	
2	indef.	early	fel.	Give me **a** green apple!	2	1	
3	def.	late	fel.	Give me **the** green apple!	1	1	
4	indef.	late	fel.	Give me **a** green apple!	2	2	
5	def.	early	infel.	Give me **the** green apple!	2	1	
6	indef.	early	infel.	Give me **a** green apple!	1	2	
7	def.	late	infel.	Give me **the** green apple!	2	2	
8	indef.	late	infel.	Give me **a** green apple!	1	1	

The actual experiment used conditions 1–6. The target in this example is always "green apple", and "green umbrella" is the competitor. (Note: determ. = determiner; def. = definite; indef. = indefinite; disamb. = disambiguation; fel. = felicitous; infel. = infelicitous; # = number of)

were to move the mouse-cursor from the start box into the corresponding response box, that is, into the corresponding shelf compartment as soon and as fast as possible.

Each trial started with the presentation of the empty shelf and the starting box. When the mouse cursor remained within the start box for 500 ms, the target and competitor objects appeared in the shelf. From then on, participants were to start their mouse movement within 5000 ms and to finish it within 12,000 ms. Stimulus presentation was initiated when the mouse cursor was moved a minimum of 60 px. outside the start box in the upper direction, within a corridor of 60 px. horizontally centered. This was done to avoid that participants leave the start box in a diagonal direction and to ensure that they were already moving the mouse when they had to make their decision. Participants were instructed to select the correct referent as soon as possible, and to pursue a smooth movement without stops and movement direction reversals.

All participants received 48 experimental trials of each of the Conditions 1–4 (see Table 2). In these trials, the determiner was always used felicitously. Participants in the *unreliable group* received additional 48 trials of Conditions 5 and 6 (although the main analyses focus on those sentences of Conditions 1–4). Thus, the unreliable group saw more items in total than the reliable group did, and as a consequence, the experiment took slightly longer. The experimental trials were divided into four blocks of 12 trials of Conditions 1–4 (for both groups) and the additional 12 trials of Conditions 5 and 6 in the unreliable group. In sum, each block comprised 48 trials for the reliable group and 72 trials for the unreliable group, presented in random order.[5] In half of the trials, the target appeared in the leftmost unit of the shelf; in the other half it appeared in the rightmost unit. Participants were instructed that they are interacting with another cooperative speaker who is in the same situation as they are, that is, who sees the same picture as the participants.

Prior to the experimental blocks, the experimenter demonstrated proper use of the mouse (i.e., without stopping or reversing the movement) with two trials from Conditions 1–4 each, that is, with eight trials in total. Subsequently, participants practiced the task on twelve trials. For the reliable group, these were three trials from Conditions 1–4; for the unreliable group, these were two trials from Conditions

[5] This is a slight deviation from the pre-registered procedure, because we decided to use unreliable items of Condition 5 and 6 equally often as the reliable items in Condition 1 and 2, and, therefore, increased the number of items in Conditions 5 and 6.

1–6. (This procedure also ensured that participants in the unreliable group already encountered unreliable conditions.)

Design and analyses

The independent variable of interest in the production task is the type of the determiner (definite vs. indefinite). We analyzed the percentage of infelicitous uses of the determiners, as well as erroneous choices of color and object separately. In addition, the total percentage of errors (i.e., infelicitous uses of determiners and wrong color/object choices) were analyzed. All comparisons were done with paired *t*-tests. Participants with more than 20% errors in total were replaced with new participants.

For the mouse-tracking task, erroneous trials (selection of the wrong object, initiation time too long, response too slow), trials with stops of the mouse movement (no movement within 200 ms) or backwards movements on the *y* axis (> 2 px.) were excluded from the data set first. Trajectories were then aligned to a common starting point ($x = 0, y = 120$). Trajectories ending in the right response box were mirrored such that all trajectories ended in the left response box. Calculation of the dependent measures AUC, MT, and TTT was based on raw trajectories (Kieslich & Henninger, 2017). To remind the reader of what these measures represent: AUC is area under the curve, MT is movement time, TTT is turn towards target.[6] To plot mean trajectories, data was time-normalized to 101 time steps. All screening and pre-processing was done by custom R-scripts and the R-package `mousetrap` (Kieslich & Henninger, 2017). For analyses of AUC, MT, and TTT, the data were screened for outliers, and trials were excluded if the respective value deviated more than 2.5 standard deviations from the design cell mean (calculated separately for each participant).

The main analyses focused on trials from Conditions 1–4, which were administered in both groups. Thus, mean AUC, MT, and TTT values were submitted to 2 × 2 × 2 mixed ANOVAs with determiner (definite vs. indefinite) and disambiguation (early vs. late) as repeated-measures, and group (reliable vs. unreliable) as a between-subject variable. Error percentages (based on selection of the wrong response box) were submitted to the same ANOVA to exclude trade-offs with speed-based measures (Liesefeld & Janczyk, 2019).

[6] In the preregistration report, we also included an analysis of an additional trajectory measure—X_{neg}—to compare reliable versus unreliable conditions. We did not include this analysis in the main text, because the observed data pattern does not match with the originally formulated hypothesis. The hypothesis and results are reported in an Online Supplement.

Table 3 Percentages of infelicitous uses and errors in the production task

	Determiner		*t*-test
	Definite	Indefinite	
Determiner	4.69	5.62	$t(59) = 0.82, p = 0.414, d = 0.11$
Color	1.67	2.50	$t(59) = 1.43, p = 0.159, d = 0.18$
Object	1.46	1.67	$t(59) = 0.44, p = 0.659, d = 0.06$
Total	6.15	7.50	$t(59) = 1.10, p = 0.274, d = 0.14$

Results

Production task

Descriptively, participants used the indefinite determiner infelicitously more often, and also made more errors for color and object choice in sentences demanding the indefinite determiner (see Table 3). However, none of the differences was significant. Participants who made more than 20% errors in total were replaced with new participants. This procedure lead to exclusion of 14 participants.

Mouse-tracking task

Mean trajectories are visualized in Fig. 4 as a function of sentence type and determiner: for the reliable group (see Panel (a)), mouse trajectories show earlier deviations into the final response locations for the early compared with the late disambiguation conditions for both types of determiners. For the unreliable group, in contrast (see Panel (b)), no such difference is readily observable. The impression thus is that participants indeed used the determiners to disambiguate sentence meaning, but only when all sentences were used felicitously, that is, in the reliable group. This impression is also reflected in AUC, MT, and TTT measures (see Fig. 5).

The ANOVA on AUC (1.57% outliers) revealed no significant main effect of group, $F(1,58) < 0.01$, $p = .974$, $\eta_p^2 < .01$. AUC was slightly larger for the indefinite determiner, $F(1,58) = 7.21$, $p = 0.009$, $\eta_p^2 = 0.11$, but determiner did not interact with either group, $F(1,58) = 0.70$, $p = 0.406$, $\eta_p^2 = 0.01$, nor disambiguation, $F(1,58) = 1.99$, $p = 0.164$, $\eta_p^2 = 0.03$. AUC was overall larger with late compared with early disambiguation, $F(1,58) = 47.72$, $p < 0.001$, $\eta_p^2 = 0.45$; however, this was mainly true for the reliable group, hence a significant interaction occurred, $F(1,58) = 65.09$, $p < .001$, $\eta_p^2 = 0.53$. The three-way interaction was not significant, $F(1,58) = 0.07$, $p = 0.789$, $\eta_p^2 < 0.01$.

Concerning MTs (1.65% outliers), all effects of the ANOVA were significant, all $F \geq 10.20$, all $ps \leq 0.002$. Because this included the three-way interaction, we ran separate 2 × 2 ANOVAs for each group. For the reliable

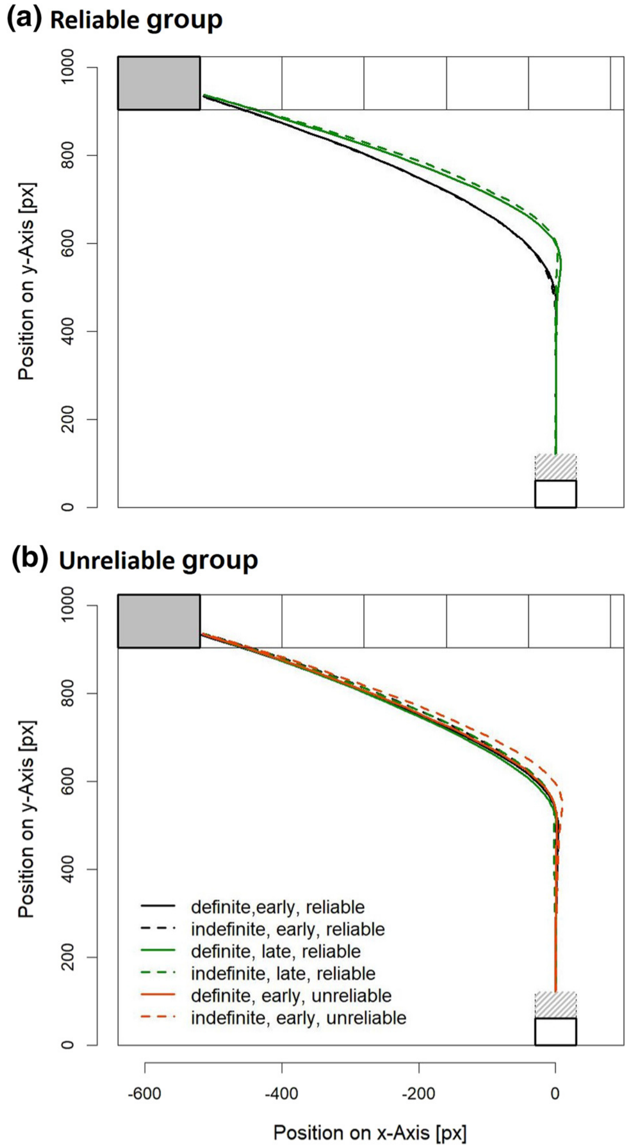


Fig. 4 Mean trajectories of the reliable group (**a**) and the unreliable group (**b**). Note that early, unreliable conditions were only implemented in the unreliable group

group, MTs were longer with late compared with early disambiguation, $F(1,29) = 85.20$, $p < 0.001$, $\eta_p^2 = 0.75$, and overall longer for the indefinite compared with the definite determiner, $F(1,29) = 68.34$, $p < 0.001$, $\eta_p^2 = 0.70$. This latter effect was, however, much more pronounced for late disambiguations; hence, a significant interaction, $F(1,29) = 57.38$, $p < 0.001$, $\eta_p^2 = 0.66$. For the unreliable group, MTs were comparable for both disambiguation conditions, $F(1,29) = 0.32$, $p = 0.577$, $\eta_p^2 = 0.01$, but longer for the indefinite compared with the definite determiner, $F(1,29) = 509.02$, $p < 0.001$, $\eta_p^2 = 0.95$. This effect was again (slightly) larger for late disambiguations; hence, a significant interaction, $F(1,29) = 8.14$, $p = 0.008$, $\eta_p^2 = 0.22$.

Regarding TTTs (3.01% outliers), only the interaction of group and determiner was not significant, $F(1,58) = 3.57$, $p = 0.064$, $\eta_p^2 = 0.06$. All other effects were significant, all $F \geq 5.01$, all $ps \leq 0.029$. Because this included the three-way interaction, we ran separate 2 × 2 ANOVAs for each group. For the reliable group, TTTs were longer for late compared with early disambiguations, $F(1,29) = 96.89$, $p < 0.001$, $\eta_p^2 = 0.77$, and overall longer for the indefinite compared with the definite determiner, $F(1,29) = 8.18$, $p = 0.008$, $\eta_p^2 = 0.22$. This latter effect was, however, only present for late disambiguations; hence, a significant interaction, $F(1,29) = 22.74$, $p < 0.001$, $\eta_p^2 = 0.44$. For the unreliable group, TTTs were slightly longer for late compared with early disambiguations, $F(1,29) = 6.57$, $p = 0.016$, $\eta_p^2 = 0.18$, and longer for the indefinite compared with the definite determiner, $F(1,29) = 32.73$, $p < 0.001$, $\eta_p^2 = 0.53$. The interaction was not significant, $F(1,29) = 1.77$, $p = 0.194$, $\eta_p^2 = 0.06$.

Error percentages are summarized in Table 4. No effect was significant, all $F \leq 2.56$, all $ps \geq 0.115$.

Discussion

We investigated whether listeners use information about uniqueness or anti-uniqueness encoded in definite and indefinite determiners to disambiguate sentence meaning as early as possible, that is, already before the noun. Participants first performed a production task in which they chose a determiner to form a contextually adequate request for one of several visually presented objects. Subsequently, they performed a mouse-tracking task where they were to select an object according to an auditorily presented stimulus sentence.

Summary of results

Error rates in the production task were overall low and participants generally used the determiners as expected. Thus, the results of the production task reveal that participants are aware of the uniqueness presupposition of the definite determiner, and that they associate the indefinite determiner with anti-uniqueness. As error rates were comparable for both determiners, choosing one or the other in forming contextually adequate object requests appears to be equally difficult.

For the reliable group, mouse-trajectories were affected by whether a sentence allowed for an early disambiguation on the determiner or not: AUC values were smaller and MTs and TTTs were shorter for the early compared to the late disambiguation condition. This suggests that participants use the information encoded in the determiner to disambiguate sentence meaning as soon as possible, that is, on the determiner. Further, the number information appears to be encoded in both determiners and is rapidly accessible. This result fits well with results from a study on the French

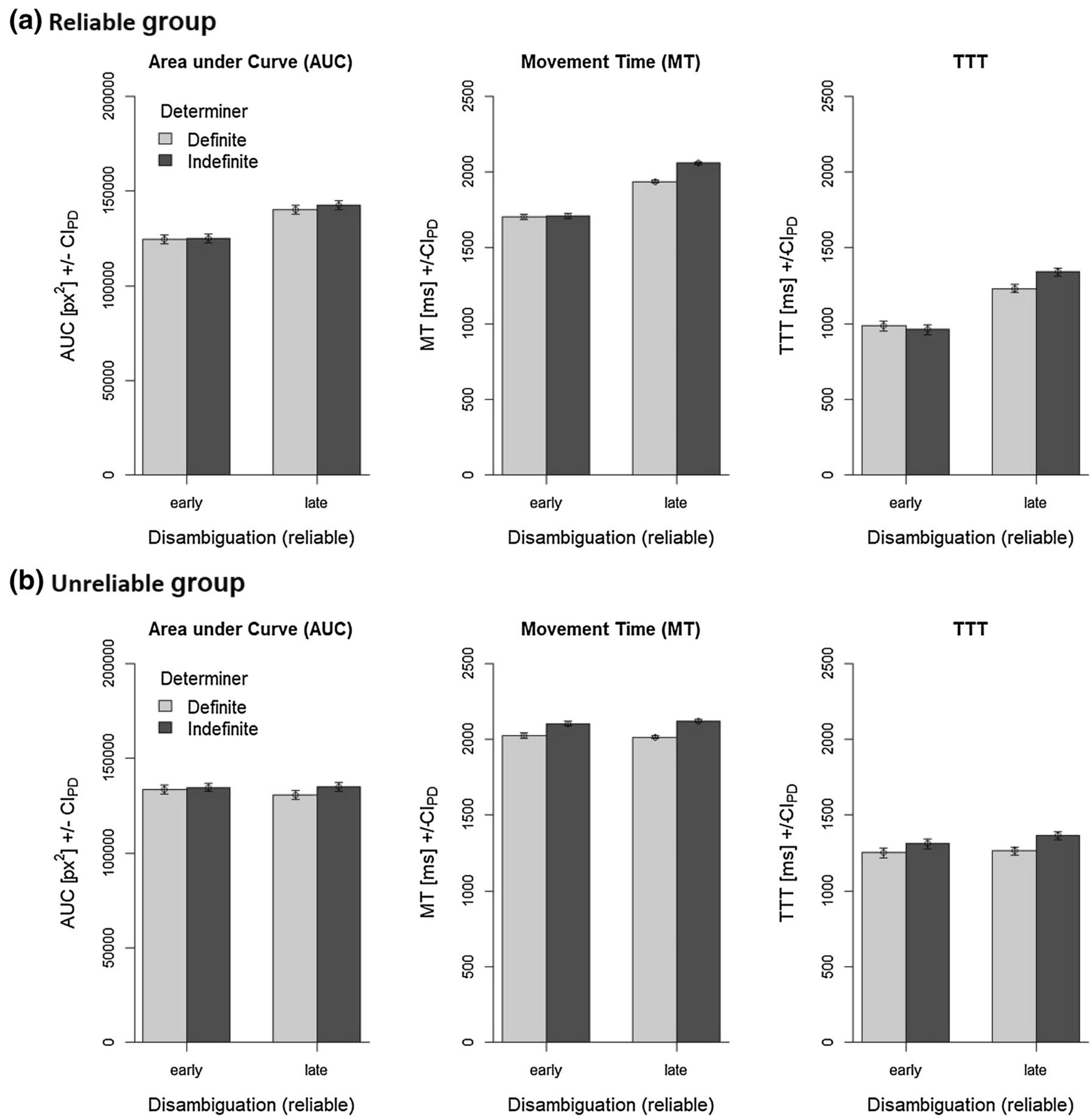


Fig. 5 Dependent measures area under curve (AUC), movement time (MT), and turn toward target (TTT) as a function of disambiguation and determiner separately for the reliable group (**a**) and the unreliable group (**b**). Error bars are 95% confidence intervals calculated separately for each comparison of the determiners (see Pfister & Janczyk, 2013)

Table 4 Error percentages in the mouse-tracking task

	Group			
	Reliable		Unreliable	
	Disambiguation		Disambiguation	
Determiner	Early	Late	Early	Late
Definite	0.50	1.06	1.41	1.12
Indefinite	1.02	0.85	1.17	1.38

language that demonstrated that gender information encoded in the determiner is also exploited immediately (Dahan et al., 2000).

However, the same was not true for the unreliable group, for which determiners were occasionally used infelicitously, and early versus late disambiguation did not affect trajectories. We conclude that exposure to infelicitous uses of determiners made participants stop using them as early cues. This suggests that deriving (anti-)uniqueness inferences must, at least in part, be a context-sensitive (pragmatic) process. In sum, the results from the mouse-tracking task yield supportive evidence for Hypotheses 1 and 3.

The results regarding Hypothesis 2 (differences between the definite and the indefinite determiner) are less clear. To start with, the data do not support Hypothesis 2c according to which the inference associated with the indefinite determiner are viewed as an implicature (Bill et al., 2018). In this case, sentences with indefinite determiners should yield less processing difficulties than sentences with definite determiners. However, no dependent variable analyzed had smaller values for the indefinite compared to the definite determiner. Considering AUC as the dependent variable, no obvious differences between the definite and the indefinite determiner are apparent, a pattern in line with Hypothesis 2a: following "the presupposition theory", both determiners come with their own uniqueness/anti-uniqueness presupposition, and we predicted no difference in processing accordingly. Note that we also observed no significant difference in error rates between determiners in the production task.

However, not all the results are consistent with Hypothesis 2a. First, for the time-based measures MT and TTT, we observed longer times for the indefinite than for the definite determiner in the late disambiguation condition. This supports Hypothesis 2b (the anti-presupposition theory) according to which the indefinite determiner causes more processing difficulties than the definite determiner due to the additional reasoning processes involved in deriving its inference. It is noteworthy that, qualitatively, we obtained the same pattern for the early disambiguation condition within the unreliable group. Thus, when the information encoded in the determiner is no longer used for immediate disambiguation, the indefinite determiner induces more processing difficulties than the definite determiner. Second, that the determiners both encode presuppositions is unlikely given the results regarding Hypothesis 3. Presuppositions should be harder to override given that they are assumed to be lexical information, whereas implicatures are (to a certain degree) optional inferences which participants can be trained to ignore (Bott & Noveck, 2004). In sum, it appears that the results are partly in line with Hypothesis 2b and thus support the anti-presupposition theory (see also Schneider et al., 2019), although we contend that some aspects of the data are in line with Hypothesis 2a as well.

Theoretical implications

According to the anti-presupposition theory, processing indefinite determiners involves an initial consideration of the uniqueness presupposition of the definite determiner (its alternative) and its subsequent negation. This should then result in more processing difficulties for the indefinite determiner compared to the definite determiner (see also Schneider et al., 2019). Why, then, was this pattern much more pronounced for the late compared to the early disambiguation condition, especially in the reliable group? One possible explanation is that in the late disambiguation condition of the reliable group, the indefinite determiner does not reduce uncertainty about which of the two non-unique objects the speaker refers to. This problem does not occur for the definite determiner because the two different objects on the shelf are unique, respectively. In the early disambiguation condition, this consideration may have played less of a role given that decisions were made already before the information from the subsequent disambiguating noun was processed. A further alternative explanation will be discussed in Sect. "Limitations and future work".

Moreover, it is in line with Hypothesis 2b that early cues are not used anymore for the indefinite determiner in the unreliable group. Given that the inference is a result of pragmatic reasoning over alternatives, it should be affected by reliability of the speaker (Rouillard & Schwarz, 2017). It remains surprising and unexplained, however, why this is not different for the definite determiner, which should be stronger in meaning according to Hypothesis 2b.

The observed pattern of performance with both definite and indefinite determiners may also partially result from particular properties of the experimental setup. From our experimental setup, it was likely clear that (1) definite and indefinite determiners are in competition, (2) the information given by the speaker was crucial to fulfill the task, and (3) the number of objects played a role. These factors combined may have led to a more parallel treatment of the two determiners than predicted by the anti-presupposition theory. In particular, the experimental setup differs from previous work in that the speaker directly addressed the participants with an order ("Gib mir...!", Engl.: "Give me ...!"). Previous experiments either asked people to evaluate statements with definite and indefinite determiners by choosing a picture fitting their interpretation (Bade & Schwarz, 2019a) or by judging sentences as true or false (Schneider et al., 2019). It was unclear in these earlier studies who the speaker was and what his/her knowledge state was. Furthermore, it was underspecified what consequences, especially for the speaker, the participants' decision had in these experiments. These factors may have contributed to the lower percentage (around 30%) to which participants actually drew the inference associated with the indefinite determiner compared to the definite determiner (92%) in Bade and Schwarz (2019a). In other words, participants did not rely on the information encoded in the indefinite determiner for their choices. The reason for this may be uncertainty about which interpretation of the indefinite determiner was intended in the experiment, given that it is ambiguous between referential and quantificational uses. This consideration may have given rise to the clear differences between the two determiners observed with eye-tracking by Bade and Schwarz (2019a). In contrast, participants in the present experiment had a clearer idea of the goals and knowledge of the speaker, and that may have driven them to use the inferences of both determiners as early as possible to predict eventual sentence meaning. Especially in the reliable group, the use of the indefinite determiner as referring to non-unique objects was very clear and unambiguous. As a result, the anti-uniqueness inference may have become more salient and thus less easily distinguishable from the presupposition of the definite determiner in processing. Further support for this view comes from a follow-up experiment of Bade and Schwarz (2019a) reported in Bade and Schwarz (2019b) which suggests that (1) target choices for the indefinite determiner are boosted by exposure to the alternative (the definite determiner) and (2) if the anti-uniqueness inferences can reliably be drawn, the processing patterns of definite and indefinite determiners do not differ.

A further factor that may have helped in stabilizing the anti-uniqueness inference associated with the indefinite determiner is the production task that preceded the mouse-tracking task. With the production task, we evaluated whether participants are aware of the (anti-)uniqueness of determiners. Participants that did not reliably use the number information were replaced. Conceivably, the production task has already shifted participants' attention to the inferences associated with both determiners and presented them as alternatives. In line with this, Bade and Schwarz (2019b) demonstrated that this affects the percentage of anti-uniqueness inferences. The authors varied the order of comprehension and production blocks between two groups. The choice of target pictures (with non-unique objects) for the indefinite determiner was much higher when the production block preceded the comprehension block than when it only followed the comprehension block. Bade and Schwarz thus hypothesized that making the alternatives explicit (with the production task) makes the anti-uniqueness inference more salient. Similar sensitivity to salient alternatives have been observed for scalar implicatures (e.g., Franke, 2014; Degen & Tanenhaus, 2015).

In sum, both (1) having very clear unambiguous uses of the indefinite determiner in the comprehension task and (2) presenting the definite determiner as an alternative in the preceding production task may have stabilized the anti-uniqueness inference and thus blurred the distinction between definite and indefinite determiners in the current experiment.

Limitations and future work

One clear limitation of the present study is that it focuses only on determiners in only one language (German). Given that determiner systems show high variability cross-linguistically, it would be unwarranted to draw very strong conclusions from our findings for the processing of (indefinite) determiners in general. This is especially the case given that the indefinite determiner in German, as in many other languages, can fulfill more than one function, for example, can refer to specific objects, can be an existential quantifier, or can be a numeral. The interpretation of the indefinite determiner cross-linguistically may also depend on what kind of indefinite competitors the language offers. German, for example, has the free choice item "irgendein" (Engl.: "anyone") as an alternative, English has "some" and the numeral "one". Given the complexity of the empirical situation, it is unclear whether all uses of the indefinite determiner (and all indefinites) are associated with the inference of interest here, and what additional factors may play a role. To test the predictions of the anti-presupposition theory, it may thus be worthwhile to test more phenomena it has been applied to (e.g., mood, tense, number, competition between "all" and "both") to avoid this additional complexity.

A potential problem for interpreting differences between determiners results from the fact that the sequence of the indefinite determiner and the following adjective is longer than the sequence of the definite determiner and the following adjective (693 ms vs. 575 ms on average). This is because the two determiners in the German accusative differ in number of letters and syllables when the noun is masculine. Additionally, the corresponding adjective differs due to declension. This may also explain differences between both determiners for the time-based measures MT and TTT in the late disambiguation conditions, where participants have to wait until the onset of the noun to make their decision. However, if this was the only explanation, one would expect this difference to be present in all conditions, including the early disambiguation condition within the reliable group: conceivably, participants need to process the whole determiner for its identification, and this should also differ in duration. Consequently, the same results would be expected as well. Admittedly, we cannot exclude that the mere onset of the determiner sufficed to distinguish both determiners in this particular condition. In this case, results should indeed be similar for both determiners. Accordingly, we should also qualify the conclusions made above and we concur that this would favor Hypothesis 2a more than Hypothesis 2b. On the basis of our data, we cannot clearly decide on this issue. One way to address this issue is to design a similar experiment in the English language, where the definite determiner "the" is not shorter compared to the indefinite determiner "a" and the adjective is morphologically equally complex. Finally, however, it should be noted that the important differences between the early and late disambiguation condition are not affected by this potential problem.

An interesting question which the present data cannot answer is whether participants learn to use cues associated with determiners during the experiment when they are exposed to reliable uses or whether participants use such cues from the very beginning, but "unlearn" to use them when confronted with infelicitous uses (i.e., in the unreliable group). It is unclear from the current experiment what role, if any, the production task played in this (un)learning. Clearly, participants' attention was already drawn to the fact that number information encoded in the determiners, as well as competition between them, may play a role. To tackle this issue, future research should consider to either not include a production task, or apply this task before or after the main experiment in two separate groups. If exposure to alternatives does play a role, this should affect the processing of determiners. Moreover, to address the issue of (un)learning, the number and order of infelicitous/felicitous uses of determiners could be manipulated in follow-up studies. If number information is encoded in the lexicon for both determiners, then a gradient unlearning effect should be observed for both of them.

Conclusion

In sum, it appears that the uniqueness and anti-uniqueness inferences associated with definite and indefinite determiners, respectively, are used rapidly to disambiguate sentence meaning and to make contingent decisions. The robustness of this is, however, affected by occasional infelicitous uses of determiners. We find that it is thus important to consider the reliability of the speaker when investigating these inferences. Regarding differences between determiners, our results are less conclusive and we refrain from drawing strong conclusions. However, the conclusion that participants used the information of the determiners to disambiguate sentence meaning as soon as possible is not undermined by this.

Supplementary material

A preregistration report including the hypotheses, the planned design of the experiment, the stimulus material, and the planned analyses was uploaded OSF prior to data collection (2018-08-03 01:53 PM; https://osf.io/aym9p/).

Acknowledgements Open Access funding provided by Projekt DEAL. This research was supported by the German Research Foundation (DFG) via Project B2 of the SFB 833 “The constitution of meaning”.

Compliance with ethical standards

Conflict of interest The authors have no competing interests or conflict of interests.

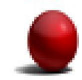
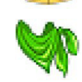
Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Appendix

Items used in the mouse-tracking experiment

Table 5 Overview of the items employed in the mouse-tracking task

item	color 1	color 2
apple		
ball		
book		
cellphone		
cup		
dress		
fur coat		
hoodie		
knife		
pencil		
raincoat		
scarf		
shirt		
sneaker		
sweater		
teddy bear		
umbrella		

Each item is depicted in its two colors

References

Alonso-Ovalle, L., Menndez-Benito, P., & Schwarz, F. (2011). Maximize presupposition and two types of definite competitors. In S. Lima, K. Mullin, & B. Smith (Eds.) *Proceedings of NELS 39* (pp. 29-40). Amherst: Department of Linguistics.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.

Bade, N., & Schwarz, F. (2019a). (In-)definites, (anti-)uniqueness, and uniqueness expectations. In R.P. Cooper (Ed.) *Proceedings of CogSci 2019* (pp. 119–125), Montreal, Canada: Cognitive Science Society.

Bade, N., & Schwarz, F. (2019b). An experimental investigation of antipresuppositions. In A. Creemers, & C. Richter (Eds.) *Proceedings of PLC 42* (pp. 29–40). Philadelphia, Pennsylvania: University of Pennsylvania Working Papers in Linguistics.

Barker, C. (1998). Partitives, double genitives and anti-uniqueness. *Natural Language & Linguistic Theory*, *16*, 679–717

Bill, C., Romoli, J., & Schwarz, F. (2018). Processing presuppositions and implicatures: Similarities and differences. *Frontiers in Communication*, *3*, 1–44.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437–457.

Chemla, E. (2008). An epistemic step for anti-presuppositions. *Journal of Semantics*, *25*, 141–173.

Chemla, E. (2009). Presuppositions of quantified sentences: Experimental data. *Natural Language Semantics*, *17*, 299–340.

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (Vol. 3, pp. 2297–2331). Berlin: Mouton de Gruyter.

Clifton, C. (2013). Situational context affects definiteness preferences: Accommodation of presuppositions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 487–501.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in french. *Journal of Memory and Language*, *42*, 465–480.

Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science*, *35*, 983–996.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicatures: A constraint-based approach. *Cognitive Science*, *39*, 667–710.

Farmer, T., Cargill, S. A., & Spivey, M. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language*, *57*, 570–595.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One*, *8*, 1–18.

Franke, M. (2014). Typical use of quantifiers: A probabilistic speaker model. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.) *Proceedings of CogSci 2014* (pp. 487–492). Austin, Texas: Cognitive Science Society.

Freeman, J. B., & Ambady, N. (2010). Mousetracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*, 226–241.

Freeman, J., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, *137*, 673–690.

Freeman, J., Dale, R., & Farmer, T. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*, 59.

Grodner, D. J., & Sedivy, J. C. (2011). The processing and acquisition of reference. In E. Gibson & N. J. Pearlmutter (Eds.), *The effect of speaker-specific information on pragmatic inferences* (pp. 239–272). Cambridge, MA: MIT Press.

Grønn, A., & Sæbø, K. J. (2012). A, the, another: A game of same and different. *Journal of Logic, Language and Information*, *21*, 75–95.

Heim, I. (1991). Artikel und Definitheit. In A. v. Stechow, & D. Wunderlich (Eds.), *Semantik /Semantics. Ein internationales Handbuch der zeitgenössischen Forschung. An international handbook of contemporary research* (pp. 487–535). Berlin: De Gruyter Mouton.

Heim, I. (2011). Definiteness and indefiniteness. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (pp. 996–1025). Berlin: De Gruyter Mouton.

Horn, L. R. (1972). *On the semantic properties of logical operators in English*. PhD thesis, University of California, Los Angeles, CA.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*, 57–83.

Janczyk, M., Pfister, R., & Kunde, W. (2013). Mice move smoothly: Irrelevant object variation affects perception, but not computer mouse actions. *Experimental Brain Research*, *231*, 97–106.

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, *49*, 1652–1667.

Kirsten, M., Tiemann, S., Seibold, V. C., Hertrich, I., Beck, S., & Beck, B. (2014). When the polar bear encounters many polar bears: Event-related potential context effects evoked by uniqueness failure. *Language, Cognition and Neuroscience*, *29*, 1147–1162.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*, 148–203.

Kratzer, A. (2005). Building resultatives. In C. Maienborn & C. Wöllenstein-Leisten (Eds.), *Event arguments in syntax, semantics, and discourse* (pp. 177–212), Tübingen, Germany: Niemeyer.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49.

Kutas, M., Petten, C. K. V., & Kluender, R. (2006). Psycholinguistics electrified. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics (Second Edition), (*pp. 659–724). London: Academic Press.

Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy tradeoffs (?). *Behavior Research Methods*, *51*, 40–60.

Magri, G. (2009). A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, *17*, 245–297.

Percus, O. (2006). Antipresuppositions. In: A. Ueyama (Ed.) Theoretical and empirical studies of reference and anaphora: Toward the establishment of generative grammar as an empirical science, Japan Society for the Promotion of Science, pp. 52–73.

Pfister, R., & Janczyk, M. (2013). Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. *Advances in Cognitive Psychology, 9,* 74-80.

Pfister, R., Janczyk, M., Wirth, R., Dignath, D., & Kunde, W. (2014). Thinking with portals: Revisiting kinematic cues to intention. *Cognition*, *133*, 464–473.

Roettger, T. B., & Stöber, M. (2017). Manual response dynamics reflect rapid integration of intonational information during reference resolution. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.) *Proceedings of CogSci 2017* (pp. 3010–3015). Austin, Texas: Cognitive Science Society.

Roettger, T. B., & Franke, M. (2019). Evidential strength of intonational cues and rational adaptation to (un-)reliable intonation. In

R. P. *Cooper Cognitive Science*, 43 (pp. e12745). Hoboken, New Jersey: Cognitive Science Society.

Rouillard, V., & Schwarz, B. (2017). Epistemic narrowing for maximize presupposition. In A. Lamont, & K. A. Tetzloff (Eds.), *Proceedings of NELS 47* (pp. 49–62). Amherst: Department of Linguistics.

Sauerland, U. (2008). Implicated presuppositions. In A. Steube (Ed.), Sentence and Context. Language, Context & Cognition., Mouton de Gruyter.

Sauerland, U., Tamura, A., Koizumi, M., & Tomlinson, J. M. (2015). Tracking down disjunction. In M. Otake, S. Kurahashi, Y. Ota, K. Satoh, & D. Bekki (Eds.), *New Frontiers in Artificial Intelligence: JSAI-isAI 2015 - Lecture Notes in Computer Science* (Vol. 10091). Springer.

Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition*, *115*, 407–416.

Schlenker, P. (2012). Maximize presupposition and gricean reasoning. *Natural Language Semantics*, *20*, 391–429.

Schneider, C., Schonard, C., Franke, M., Jäger, G., & Janczyk, M. (2019). Pragmatic processing: An investigation of the (anti-)presupposition of determiners using mouse-tracking. *Cognition*, *193*, 104024.

Schneider, C., Bade, N., & Janczyk, M. (2020). Is immediate processing of presupposition triggers automatic or capacity-limited? A combination of the PRP approach with a self-paced reading task. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/s10936-019-09686-3

Schwarz, F., & Tiemann, S. (2012). Presupposition processing: The case of German wieder. *Language and Meaning, 7218*, 200–209.

Schwarz, F. (2007). Processing presupposed content. *Journal of Semantics*, *24*, 373–416.

Schwarz, F. (2016). Experimental work in presupposition and presupposition projection. *Annual Review of Linguistics*, *2*, 273–292.

Singh, R. (2011). Maximize presupposition! and local contexts. *Natural Language Semantics*, *19*, 149–168.

Spivey, M. (2008). *The continuity of mind*. Oxford: University Press.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 10393–10398.

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*, 227–267.

Stranahan, L. (2018). *Effects of working memory load and speaker reliability on contrastive inference and quantifier processing*. PhD thesis, Harvard University.

Tiemann, S. (2014). *The processing of wieder ('again') and other presupposition triggers.* PhD thesis, University of Tuebingen.

Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., et al. (2011). Psycholinguistic evidence for presuppositions: on-line and off-line data. In I. Reich, E. Horch, & D. Pauly (Eds.) *Proceedings of Sinn & Bedeutung 15* (pp. 581–595). Saarbrücken: Saarland University Press.

Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, *69*, 18–35.

van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, *41*, 147–182.

van Berkum, J., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology*, *40*, 235–248.

Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (2016). Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language*, *87*, 128–143.

10.5 Study 5

Schneider, C., Janczyk, M., & Bade, N. (under review). Verifying pragmatic content in context: An experimental comparison of presuppositions of *again* and the definite determiner with scalar implicatures. *Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data.* (submitted on 26.05.2020)

Verifying pragmatic inferences in context: An experimental comparison of presuppositions of *again* and the definite determiner with scalar implicatures

[a]Cosima Schneider(ORCID: 0000-0002-4314-0885), [b]Markus Janczyk (ORCID: 0000-0002-9958-3220), & [c]Nadine Bade

[a]University of Tübingen, Tübingen, Germany
[b]University of Bremen, Bremen, Germany
[c]Ecole Normale Supérieure (ENS), Paris, France

cosima.schneider@uni-tuebingen.de

1 Introduction

The present paper investigates a fundamental issue of presupposition (PSP) processing, that is, the validation of PSPs in a given context. To that end, we compare test sentences that trigger PSPs in contexts that either falsify or verify them. We focus on two particular triggers, namely the definite determiner "the" (German *der/die/das*) and "again" (German *wieder*).

Overall, we pursue four goals. First, we aim to replicate results indicating an early processing of PSPs starting as soon as the presuppositional content is fully known (Tiemann et al., 2011) with a larger sample size. Second, we aim to investigate classifications of triggers offered in the theoretical literature by comparing two PSP triggers assumed to belong to different classes. Third, we investigate how scalar implicatures (SIs) are processed, and fourth, we compare the time-course of processing PSPs and SIs in verifying/falsifying contexts, to further explore the current theoretical distinction of different types of pragmatic content more generally.

2 Theoretical approaches to Presuppositions and Implicatures

2.1 Presuppositions

Theoretical approaches to PSPs. According to a traditional **semantic approach**, PSPs are encoded in the lexical entry of a PSP trigger, that is, in a lexical item. PSPs are regarded as "definedness conditions": a sentence *s* presupposes *p* just in case *p* must be true in order for *s* to have a truth value at all (Strawson, 1952). If the PSP is not met, the sentence is undefined rather than true or false (Heim & Kratzer, 1998). Formally, this is captured by making sentences with PSPs denote partial functions, and a sentence like (1) is considered true if Joan scored and scored before, false if she did not score but scored before, and undefined otherwise.

(1) Joan scored, again.

Semantic undefinedness is mapped onto pragmatic inappropriateness through "Stalnaker's Bridge" (von Fintel, 2004).

At least some PSPs have been argued to have a pragmatic source (Abusch, 2002, 2010; Schlenker, 2010; Simons, 2001; Stalnaker, 1973) (=**pragmatic approach**), that is, they are not conventionally (=lexically) encoded but are derived contextually. Based on different triggering mechanisms, Abusch (2002) introduced the distinction of *hard* and *soft* PSP triggers. The two types behave differently in contexts where the speaker is ignorant with regard to the truth of the PSP. For example, soft triggers can be uttered in explicit ignorance contexts[1] without leading to infelicity, but hard triggers cannot be. Examples for *soft* triggers are aspectual verbs as *stop* and *continue* (see Simons, 2001), *know*, and achievement verbs as *win*. Hard triggers are, for example, *too, also, even, again,* the negative polarity item *either*, and *it-clefts* (see also Abusch, 2010). *Soft* triggers share certain properties like suspendability and non-detachability, which are typical of conversational inferences like implicatures. Based on that observation, Romoli (2011) proposed that PSPs of soft triggers *are* SIs, which come out as entailments in certain

[1]Simons (2001) introduced the term *explicit ignorance contexts* to describe situations in which it is apparent to the addressee that the speaker is ignorant with respect to the proposition that would normally be presupposed.

environments. It is still under debate which trigger belongs to which class, and the resulting effects discussed in this regard are quite subtle.

From both the pragmatic and semantic approach one can derive the hypothesis that the listener is confronted with increased cognitive workload in cases where the PSP is false or unfulfilled. However, both approaches diverge in their predictions about **when** this increase becomes apparent. According to the semantic approach, the PSP is encoded in the trigger. Difficulties are expected to arise upon encountering the trigger and should persist until the content of the PSP is known. According to the pragmatic approach, the sentence is first interpreted compositionally, and only then is the context checked for whether it is appropriate given its PSP. The distinction need not be this clear cut though, and it is possible to assume that pragmatic information is taken into consideration at every given point during computation, for example, with the assumption of local contexts (see Schlenker, 2011).

Experimental evidence. Recent years came with an increasing amount of experimental studies on PSP processing (see Schwarz, 2016, for a review). Results from some studies support the semantic approach. For example, Tiemann et al. (2011) reported effects on the trigger, when comparing presuppositional sentences to plainly grammatical and ungrammatical ones (see also Schneider, Bade, & Janczyk, 2020). Further, Tiemann (2014) and Tiemann et al. (2015) discussed processing difficulties on the trigger when the PSP is not met in the context. Similar results are discussed for reading time (RT) studies on German *auch* (Engl. *also*; Schwarz, 2007), and the definite determiner (van Berkum et al., 1999; van Berkum et al., 2003). Early or immediate processing of the uniqueness PSP of the definite determiner has furthermore been suggested by EEG studies comparing definite and indefinite determiners (Kirsten et al., 2014), as well as mouse- and eye-tracking studies (Bade & Schwarz, 2019b; Schneider, Bade, Franke, & Janczyk, 2020; Schneider, Schonard, Franke, Jäger, & Janczyk, 2019).

While many of the aforementioned studies focused on one or two PSP triggers, some studies also suggest that different types of PSP triggers may differ in processing (Abrusan, 2011; Domaneschi et al., 2013; Domaneschi & Di Paola, 2018), especially in the verbal domain. Along these line, Tiemann (2014) discussed *the* and *again* as falling into two different classes, especially regarding accommodation.

2.2 Scalar implicatures

Another prominent example of pragmatic inferences are SIs, which are theoretically distinguished from PSPs based on their properties. An example is given in (2).

(2) Some elephants are mammals.

SI: Not all elephants are mammals.

Theoretical approaches. Under a **Gricean** (or pragmatic) **view** (Grice 1989), implicatures arise as a result of pragmatic reasoning based on the *Cooperative Principle* and the four Gricean maxims of conversation. The SI in (2) is the result of reasoning with quantity: as the sentence with *all* is strictly more informative, the speaker deduces that the hearer would have used it if s/he believed it to be true. As s/he did not, the hearer can safely assume that the speaker does not believe the *all* alternative to be true. According to Grice, the process is highly context-dependent, and the resulting inferences are weak and cancelable.[2]

Under a **lexical approach** (Chierchia, 2004; Levinson 2000), a lexical ambiguity is essentially argued for. The implicature is added to the lexical meaning of the scalar term in such a way that *some* has a meaning *some and not all*, which is claimed to be more efficient than assuming reasoning over alternatives in several steps. Levinson (2000) proposed that the SI is marked as defeasible in the lexicon, whereas Chierchia (2004) slightly deviated from Levinson

---

[2] Under so called Neo-Gricean approaches (Sauerland, 2004), the basic idea of SIs as the result of pragmatic reasoning is kept, but a strengthening mechanism is introduced, as well as the notion of local implicatures. We won't discuss this approach in much detail here, as it is not relevant for the predictions we tested.

and introduced a scalar term with a *weak* (implicature-free) meaning and with a *strong* (implicature-laden) meaning.

To explain how listeners arrive at an interpretation of the *pragmatic some* (*some, but not all*), Huang and Snedeker (2009) introduced a processing model based on the Gricean and lexical approach. This **Literal-First view** assumes that the lower-bound semantic interpretation (*some and possibly all*) is computed rapidly as part of the basic sentence meaning. All inferences including an SI and the upper-bound meaning require extra time and resources, and processing takes place in two steps: First, the semantic meaning (*some and possibly all*) is constructed, and second, the pragmatic meaning (*some, but not all*) is computed. Two steps are necessary, because the pragmatic interpretation cannot exist without the semantic one. Accordingly, the pragmatic (or upper-bound) meaning of *some* requires more processing resources and thus takes longer than the processing of the semantic (or lower bound) meaning of *some*.

Finally, according to the **grammatical approach,** SIs arise as entailments of exhaustified sentences (see, e.g., Chierchia, 2006; Fox, 2007; Magri, 2010). The exhaustivity operator EXH that derives the exhaustification is similar in meaning to overt *only* (modulo presuppositions). It applies to a proposition *p* and its alternatives and affirms the proposition while negating the subset of relevant alternatives.

Processing of SIs under this **Default view** (Levinson, 2000) is predicted to be effortless and immediate, because of their status as default inference, and only cancellation of the implicature incurs processing recourses. This view is motivated by the articulatory bottleneck, which claims that communication proceeds remarkably quickly although humans can only produce a highly limited number of phonemes per second. It predicts that the upper-bound interpretation SI (i.e., *some, but not all*) precedes the lower-bound interpretation (*some, and possibly all*) and thus the semantic (or lower bound) meaning of *some* should require more processing resources and take longer than processing of the pragmatic (or upper-bound) meaning of *some.*

Experimental evidence. The empirical picture on implicatures is complex, as they have been shown to be influenced by a variety of factors and have been investigated extensively with a variety of methods. Due to space limitations, we cannot discuss the experimental literature in detail here. An overview of the historical development of the experimental literature on implicatures can be found in Noveck (2018). Broadly speaking, implicature derivation has been shown to be more costly than accessing literal meaning, supporting a literal-first view. Results diverge with regard to when effects occur. Some studies reported delayed processing of implicatures associated with *or* and *some* (e.g., Bott & Noveck, 2004; Bott et al., 2012; Breheny et al., 2006; Chevallier et al., 2008; Huang & Snedeker, 2009, 2011; Noveck & Posada, 2003). However, other experimental data, for example from eye tracking (Breheny et al., 2013; Foppolo & Marelli, 2017, Grodner et al., 2010), suggest that implicatures are immediately available (see also Sedivy et al., 1999). Crucially for our purposes, differences in processing between PSPs and SIs seem to confirm their different theoretical status (Bill et al., 2018).

3 The Experiment

Our experiment is a conceptual replication of Tiemann et al.'s (2011) Experiment 2. As these authors, we employed self-paced reading to investigate at which point in time a PSP verification with the context takes place. Tiemann et al. included five different PSP triggers in their experiment (German *wieder*, Engl. *again*; *auch*, Engl. *also*; *aufhören*, Engl. *stop*; *wissen*, Engl. *know*; and definites in the shape of possessive noun phrases [German *sein/ihr,* Engl. *his/her*]), and compared sentences in contexts that either (i) verified the PSP or (ii) falsified it. RTs at the positions of the trigger, the word following the trigger, and the evaluation word (the word where the content of the PSP is known) were analyzed. The results supported the view that a validation process of the PSP starts as soon as the PSP is known, and before the end of the sentence. In a falsifying context this validation process took longer compared to a context that verified the PSP. However, because recent studies suggested that different types of PSP triggers may differ in processing (Abrusan, 2011; Domaneschi et al., 2013; Domaneschi & Di Paola, 2018; Tiemann et al., 2015), it is unfortunate that Tiemann et al. (2011) did not report separate analyses of PSP triggers.

Here, we focus on two PSP triggers, the definite determiner *the* (German *der/die/das*) and its uniqueness-PSP and *again* (German *wieder*). This focus allows for using a larger number of items per participants and condition and, as a result, meaningful analyses of differences between the two PSP triggers. We additionally included SIs to contrast their processing with that of PSPs. Sentences with scalar *some* (German *einige*) were used for the implicature condition, and sentences with *all* (German *alle)* were included as a control. They should not evoke an implicature as they are already most informative, and can only be either literally true or false.

3.1 General approach and procedure

For each sentence type (determiner, again, SI), we created 40 sets of experimental sentences, thus 120 sets in total, with two different context sentences each (see (A) and (B)). Each context sentence was paired with two test sentences, in such a way that the context sentence either verified the PSP/SI of the test sentence or falsified it. If context (A) verified the content of the PSP/SI, then context (B) falsified it and vice versa. We tested participants in one session (in contrast to Tiemann et al., 2011). Participants saw each item in only one context condition (A or B), but saw both test sentences for that context, one falsified and one verified by the context (see Table 1 for example items).

To facilitate the comparison of the different regions of interest we refer to the words of interest via the region they appear in: P1 is the "trigger word" (scalar term for SIs and lexical trigger for PSPs; underlined in Table 1). P2 is the critical word, that is, the point when the complete content of inferences was known, and P3 is the final word. The additions +1/+2 refer to the words following P1, P2, and P2, respectively (see also Table 2 in the Appendix).

Table 1. Example items with the two context variations in the verifying and falsifying condition.

Example item: determiner					
(A)	Manuel hat ein Ticket für ein Baseball-spiel gekauft.		(B)	Manuel hat mehrere Tickets für ein Base-ballspiel gekauft.	
	Manuel bought a ticket for a baseball match.			*Manuel bought several tickets for a base-ball match.*	
	1	Manuel holt das Ticket und freut sich.		3	Manuel holt die Tickets und freut sich.
		Manuel collects the ticket and he is happy.			*Manuel collects the tickets and he is happy.*
	2	Manuel holt die Tickets und freut sich.		4	Manuel holt das Ticket und freut sich.
		Manuel collects the tickets and he is happy.			Manuel holt das Ticket und freut sich.
Example item: again					
(A)	Lukas hat schon oft Pizza bestellt.		(B)	Lukas hat noch nie Pizza bestellt.	
	Lukas has often ordered pizza before.			*Lukas has never ordered pizza before.*	
	5	Heute hat Lukas wieder Pizza bestellt und wartet freudig.		7	Heute hat Lukas wieder keine Pizza bestellt und hat nichts zu essen.
		Today, Lukas ordered pizza again and waits happily.			*Today, Lukas didn't ordered pizza again and has nothing to eat.*
	6	Heute hat Lukas wieder keine Pizza bestellt und hat nichts zu essen.		4	Heute hat Lukas wieder Pizza bestellt und wartet freudig.
		Today, Lukas didn't ordered pizza again and has nothing to eat.			*Today, Lukas ordered pizza again and waits happily.*
Example item: SI					
(A)	Zwei von vier Schrauben sind kaputt.		(B)	Vier von vier Schrauben sind kaputt.	
	Two of four screws are broken.			*Four of the four screws are broken.*	
	9	Weil einige Schrauben kaputt sind müssen neue gekauft werden.		11	Weil alle Schrauben kaputt sind müs-sen neue gekauft werden.
		Because some of the screws are bro-ken they have to buy new ones.			*Because all of the screws are broken they have to buy new ones.*
	10	Weil alle Schrauben kaputt sind müssen neue gekauft werden.		12	Weil einige Schrauben kaputt sind müssen neue gekauft werden.
		Because all of the screws are broken they have to buy new ones.			*Because some of the screws are broken they have to buy new ones.*

3.2 Purpose of the present study and hypotheses

The present study pursued four major goals. First, we aimed to replicate the results of Tiemann et al.'s (2011) Experiment 2. We therefore expect that falsification of a PSP causes higher processing difficulties than its verification, being reflected by longer RTs in the falsification than in the verification condition (see also Altmann & Steedman, 1988; Hertrich et al., 2015; Kirsten et al., 2014; Schwarz, 2007) (=Hypothesis 1, H1). These difficulties are expected in different regions depending on the process being semantic (early effects) or pragmatic (late effects).

Second, we aimed at comparing processing of Class 1 with the Class 2 PSP triggers (see Glanzberg, 2005, or Tiemann et al., 2015, for this distinction) and Hypothesis 2 (H2) states to observe differences in processing of the two PSP triggers (see also Abrusán, 2011; Domaneschi et al. 2014; Domaneschi et al., 2018; Domaneschi & Di Paola, 2018; Jouravlev et al., 2016; Tiemann et al., 2015). For example, the effect of context condition might be larger for definite determiners than for *again*, because in the latter sentences, the PSP can be ignored without making the sentence entirely senseless, that is, assertion and PSP are not dependent on each other (see Tiemann, 2014, and Tiemann et al., 2015) for a theoretical discussion and experimental data).

Third, we explored processing of SIs in comparison to PSPs. For SIs, processing difficulties are expected (H3) and, depending on the theoretical approach, these difficulties appear in different conditions: we expect longer RTs in the falsifying condition under the Default view, but in the verifying condition under the Literal-First view. Furthermore, we expect differences between PSPs and SIs. Hypothesis 4 (H4) reflects our expectation of early effects starting on the trigger for the PSPs (see, e.g., Bade & Schwarz, 2019a; Burkhardt, 2006; Kirsten et al., 2014; Schneider et al., 2020; Schwarz, 2007; Schwarz & Tiemann, 2012; Tiemann et al., 2011), while for SIs slower and/or delayed processing is expected (see e.g., Bott & Noveck, 2004; Chevallier et al., 2008; Chierchia et al., 2001; De Neys & Schaeken, 2007; Feeney et al., 2004; Huang & Snedeker, 2009; Noveck & Posada, 2003; Pouscoulous, Noveck, Politzer, & Bastide, 2007).

4 Method

4.1 Participants

The intended sample size in this experiment was *n* = 48 native speakers of German. Data were collected from 52 participants from the Tübingen (Germany) area. Two participants were excluded because of 30% or more errors in the final comprehension questions (see below). Another participant was excluded because German was not his/her mother tongue, and one participant was excluded due to technical problems during the experiment (final sample: mean age = 23.1 years, 39 females, 9 males). Participants signed informed consent prior to data collection and were paid 8€ or received course credit for participation.

4.2 Apparatus and Stimuli

Stimulus presentation and response collection were controlled by a standard PC connected to a 17-inch CRT monitor. A trial started with a context sentence presented as a whole in the upper half of the screen. After reading the context sentence, participants requested the test sentence with a button press of the right index finger on an external response button placed to the right of them. The test sentence was presented word-by-word in a self-paced reading manner, that is, the letters of the test sentences' words were first substituted by underscores as placeholders. Pressing the external key revealed the first word. To continue reading, participants had to press the button again to reveal the next word while the previous word disappeared and was again substituted with underscores. For *again* sentences, we presented *again not* (*wieder keine*) together to facilitate a comparison of the respective regions. Effects of word length did not play a role because we analyzed RT per letter (see Schneider et al., 2020; Tiemann et al., 2011). After participants finished reading the test sentence, they had to rate its appropriateness according to the presented context via the number keys 1-4 on a standard QWERTZ keyboard ranging from very unnatural (1) to very natural (4). Participants were asked to pay attention to the content, because after the experiment they had to answer 10 yes/no comprehension questions (on the basis of which two participants were excluded).

Participants started with reading instructions. This was followed by a short practice block of 24 trials, four of each sentence type in both conditions. The order of these practice trials was determined randomly, but then kept constant for all participants. Subsequently, the 240 test trials were administered in three blocks of 80 trials each. The trials were presented in random order with the restriction that sentences of the same item did not appear in different conditions in direct succession. All participants were tested individually in a single session of about 45 minutes.

4.3 Design and Analyses

The independent variables of interest were (1) context condition (verifying, falsifying) and (2) sentence type (determiner, *again*, *some*, *all*). Mean acceptability ratings were submitted to a 2 × 4 Analysis of Variance (ANOVA) with context condition and sentence type as repeated-measures.

There are three regions of interest (see also Table 2 in the appendix): The position of the PSP trigger or the scalar term (Position 1; P1), the word of evaluation, which is the word where the content of the PSP or the SI could be evaluated (Position 2; P2), and the final word of the sentence (Position 3, P3). To be able to uncover spillover effects, we additionally analyzed one word following the trigger/scalar term (P1+1), one word following the evaluation word (P2+1), and two words following the evaluation word (P2+2). In case of the determiner, the word following the PSP trigger (P1+1) is already the evaluation word (P2) (except for two items which were excluded from the analysis). Therefore, we used the same data in the ANOVA for both positions. Trials in which RTs deviated more than 2.5 SDs from the respective design cell in any of the analyzed positions (calculated separately for each participants) were excluded as outliers (15.0% of the trials). For each region, mean RTs were submitted to the same ANOVA as acceptability ratings. In case of a significant interaction, we analyzed the sentence types separately with follow-up ANOVAs with context condition as a repeated measure.

5 Results

5.1 Acceptability rating

Results of the acceptability ratings are visualized in Figure 1. Verifying conditions were clearly rated better than falsifying conditions for all four sentence types, $F(1, 47) = 676.46$, $p < .001$, $\eta_p^2 = .94$. The main effect of sentence type was also significant, $F(3, 141) = 27.74$, $p < .001$, $\eta_p^2 = .37$, reflecting the slight differences between the sentence types. The high ratings in the verifying condition suggest that participants perceive them as appropriate. In contrast, the falsifying condition received low ratings which indicates that inappropriateness was detected. The interaction was also significant, $F(3, 141) = 31.28$, $p < .001$, $\eta_p^2 = .40$, but the effect of context condition was significant for all sentence types, all *p*s < .001.

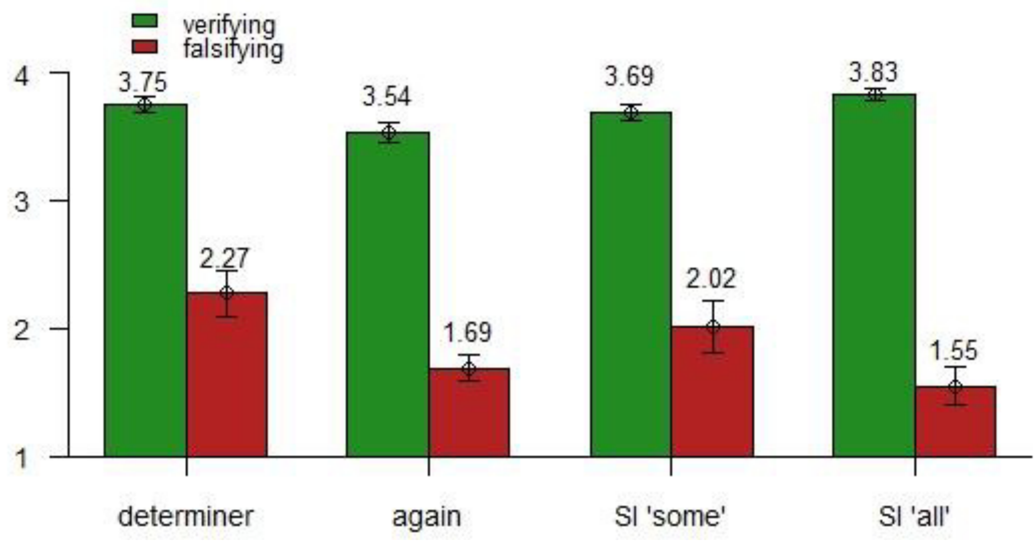


Figure 1. Mean acceptability ratings in as a function of context condition and sentence type. Error bars are 95% confidence intervals for the means (SI = scalar implicature)

5.2 Reading times

RTs per letter are visualized in Figure 2 for all sentence types. Details on inferential statistics are summarized in Table 3 in the Appendix. In short, the ANOVAs revealed significant differences between the four sentence types for all analyzed positions, indicating that different processing difficulties were evoked by these sentence types.

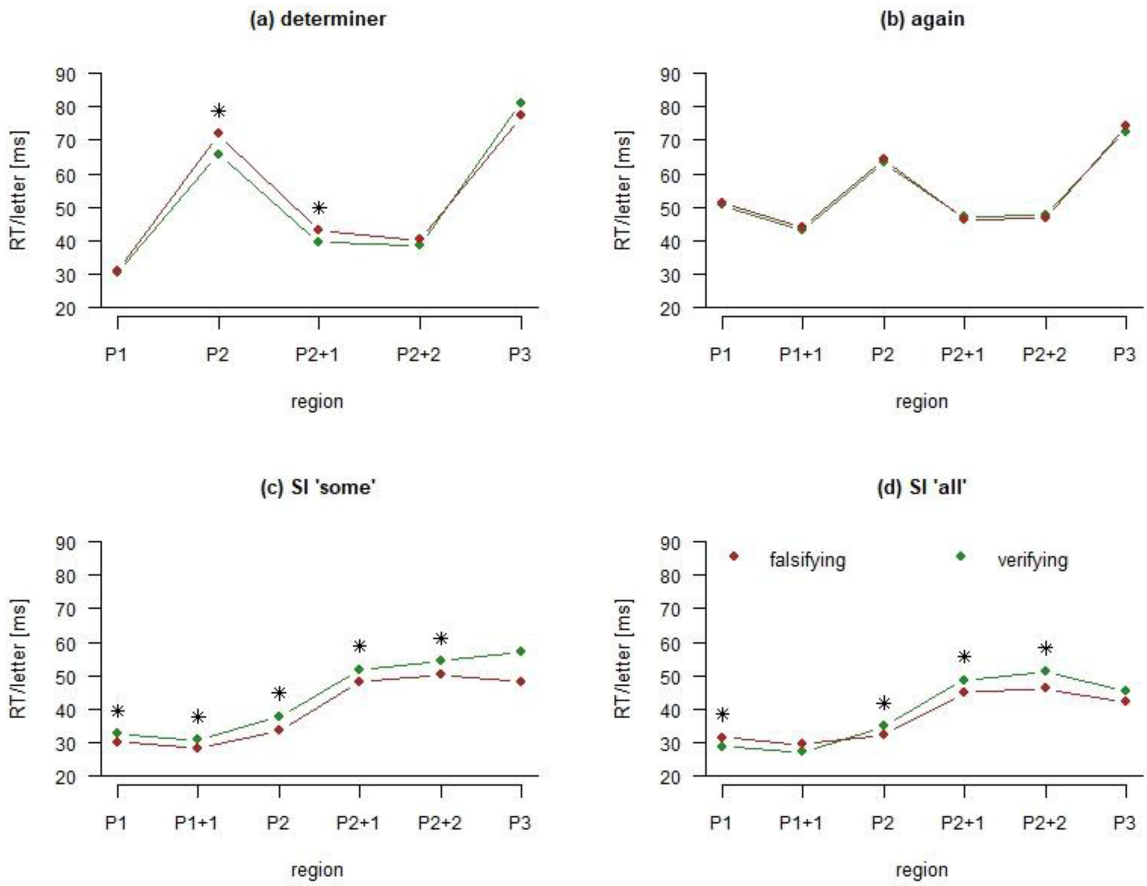


Figure 2. Reading times (RT; in milliseconds per letter) separately for the four sentence types determiner in (a), *again* in (b), implicatures with *some* in (c) and for true/false assertion with *all* in (d) for the respective regions. The asterisk marks significant differences between the sentence conditions (*P1* reflecting the trigger/scalar term, *P1+1* the word following the trigger/scalar term, *P2* the word of evaluation where the PSP/SI can be evaluated, *P2+1* one word following the word of evaluation, *P2+2* two words following the word of evaluation, *P3* the final word of the sentence, SI = scalar implicature).

For the PSP trigger/scalar term (P1), the main effect of context condition was not significant. However, we observed a significant interaction of context condition and sentence type. Context condition only had a significant impact for *some* sentences (with longer RTs in the verifying context condition) and *all* sentences (with longer RTs in the falsifying context condition), but not for the two PSP triggers.

For the word following the PSP trigger/scalar term (P1+1), additionally the main effect of context condition and the interaction were significant. The effect of context condition was only significant for determiner sentences (longer RTs in falsifying contexts) and for *some* sentences (longer RTs in verifying contexts).

At the evaluation word (P2), the main effect of context condition was not significant, but the interaction was. Context condition had an effect for determiners (longer RTs in falsifying contexts) and for *some* and *all* sentences (longer RTs in verifying contexts)

For the word following the evaluation word (P2+1) the main effect of context condition and the interaction were significant as well. Effects of context condition were observed for determiner (with longer RTs in the falsifying condition) and for *some* and *all* sentences (with longer RTs in the verifying condition). For *all* sentences this could reflect the point where the sentence stopped to make sense to the participants (i.e., where they know the assertion is false).

For the next word of interest (P2+2), the main effects of context condition and the interaction were significant. Context condition only had a significant impact on *some* and *all* sentences (with longer RTs in the verifying condition).

For the final word (P3), neither the main effect of context condition nor the interaction were significant. Descriptively, RTs for *again* were longer in the falsifying than in the verifying condition, while the opposite was true for the other three sentence types (with a slightly larger difference for *some* and *all* sentences as compared to determiner sentences).

6 Discussion

The results of the acceptability rating replicate those of Tiemann et al. (2011) by and large. The ratings for all sentence types are high in the verifying condition and low in the falsifying condition. Accordingly, when the context does not verify the PSP/SI, it is perceived as inappropriate or unnatural. The SI *all* sentences can be regarded as a baseline, because in these sentences verifying/falsifying means literally true vs. false. Thus, without requiring pragmatic enrichment, the verifying condition was perceived as natural.

RT results are only partly in line with Tiemann et al.'s (2011) observations. For the trigger position, we observed the same pattern as in the original study, namely that RTs did not differ between the two context conditions. Regarding performance on the evaluation word, our observations differed from Tiemann et al.'s results. For determiner sentences, we observe longer RTs in the falsifying condition than in the verifying condition. This result suggests an immediate verification process of the PSP as soon as the content of the PSP is fully known and thus supports the semantic approach to PSPs (H1). If the content of the PSP is not supported by the context, this process fails, leading to processing difficulties which become reflected in longer RTs at the evaluation word (but not at the end of the sentence). Thus, the data provide evidence for the semantic approach to PSPs triggered by definite determiners and rather is not in line with the pragmatic approach.[3] However, we do not see this difference for *again* sentences. This difference highlights the necessity to analyze different PSP triggers separately, just as is suggested by the classifications of Tiemann et al. (2015) and Glanzberg (2005). As the processing of the two triggers under investigation here differs, the data are in line with H2.

For SIs there is a significant difference between the two context conditions already at the position of the scalar term. The longer RTs in the verifying condition for the *some* sentences

[3] At least one that does not assume local contexts to play a role for pragmatic processing/context integration.

support the idea that SIs are only calculated in context conditions that support the implicature (see also Breheny, Katsos, & Williams, 2006; Hartshore, 2015). Regarding H3, the data thus provide evidence for the Literal-First view: sentences with the weaker scalar term led to processing difficulties in the verifying condition, when the implicature had to be calculated, not in the falsifying condition, where no implicature needed to be derived. Furthermore, contrary to previous work by Noveck and Posada (2003) and Huang and Snedeker (2011), this process starts quite early, and thus is in line with work by Grodner et al. (2010), Breheny et al. (2013) and Foppolo and Marelli (2017) who suggested that implicatures are immediately available.

Last, we can see that processing of SIs differs from processing PSPs (H4), because for SIs the effect of context condition persists longer, until the end of the sentence. Thus, in contrast to PSPs the evaluation process of SIs appears to be a long-lasting process.

There are two unexpected and interesting results. First, the reversed effect of context condition for the *all* sentences requires attention. Tentatively, we suggest the following explanation: The longer RTs for the falsifying context at the beginning of the sentence suggest that false assertions are detected rapidly by participants. This leads to processing difficulties at the scalar term and at the following word. At that point, participants realized that the sentence stopped making sense and consequently they did not process the assertion in the falsifying condition properly. This could then explain the shorter RTs in the falsifying condition in the later parts of the sentence.

Second, we did not observe a strong effect of context for the trigger *again,* contrary to previous findings. This may be either due to the material we used, which is possibly too unspecific with regard to the truth of the PSPs. It may also be due to the very complex interaction of *again* with negation, which makes verification and falsification more complex, (see Schwarz & Tiemann, 2017 for more discussion). We have to leave it to further research to address these issues in more detail.

References

Abrusán, M. (2011) Predicting the presuppositions of soft triggers. *Linguistics and Philosophy, 34,* 491-535.

Abusch, D. (2002). Lexical alternatives as a source of pragmatic presupposition. In Brendan Jackson (Ed.), *Semantics and Linguistic Theory (SALT*) 12 (pp. 1-19), California: University of California.

Abusch, D. (2010). Presupposition triggering from alternatives. *Journal of Semantics 27*, 37-80.

Altmann, G., & Steedman, M. (1988). Interaction with context during sentence processing. *Cognition, 30*, 191–238.

Bade, N., & Schwarz, F. (2019a). (in-)definites, (anti-)uniqueness, and uniqueness expectations. In R.P. Cooper (Ed.) *Proceedings of CogSci 2019* (pp. 119–125), Montreal, Canada: Cognitive Science

Bade, N. & Schwarz, F. (2019b). An experimental investigation of antipresuppositions. In A. Creemers, & C. Richter (Eds.) *Proceedings of PLC 42* (pp. 29–40). Philadelphia, Pennsylvania: University of Pennsylvania Working Papers in Linguistics.

Society.Bade, N., & Schwarz, F. (2019b). An experimental investigation of antipresuppositions. In A. Creemers, C. Richter (Eds.) *Proceedings of PLC 42* (pp. 29-40). Philadelphia, Pennsylvania: University of Pennsylvania Working Papers in Linguistics.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437-457.

Bott, L., Bailey, T. M., & Grodner, D. (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language, 66,* 123–142.

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes, 28,* 443–467.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*, 434-63.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language, 98*, 159–168.

Chevallier, C., Noveck, I. A., Nazir, T., Bott, L., Lanzetti, V. & Sperber, D. (2008). Making disjunctions exclusive. *Quarterly Journal of Experimental Psychology, 61*, 1741-1760.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Belletti structures and beyond* (pp. 39-103). Oxford: Oxford University Press.

Chierchia, G., (2006). Broaden your views: Implicatures of domain widening and the ''logicality'' of language. *Linguistic Inquiry 37*. 535–90.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Luisa, M. (2001). The acquisition of disjunction: evidence for a grammatical view of scalar implicatures. In A. H.-J. Do, L. Domínguez, & A. Johansen (Eds.) *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 157-168). Boston: Cascadilla Press.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: dual task impact on scalar implicature. *Experimental Psychology*, *54*, 128-33.

Domaneschi, F., & DiPaola, S. (2018). The processing costs of presupposition accommodation. *Journal of Psycholinguistic Research*, *47,* 483-503.

Domaneschi, F., Carrea, E., Penco, C., & Greco, A. (2014). The cognitive load of presupposition triggers: Mandatory and optional repairs in presupposition failure. *Language, Cognition and Neuroscience, 29,* 136-146.

Feeney, A., Scrafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology*, *58*, 121-32.

Foppolo, F., & Marelli, M. (2017). No delay for some inferences. *Journal of Semantics, 34,* 659–681.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In U. Sauerland & P. Stateva (Eds.), *Presupposition and implicature in compositional semantics,* (pp. 71–120). UK: Palgrave Macmillan.

Glanzberg, M. (2005). Presuppositions, truth values and expressing propositions. In G. Preyer and G. Peter (Eds.), *Contextualism in philosophy: Knowledge, meaning, and truth* (pp. 349-396). Oxford, UK: University Press.

Grice, P. (1989). *Studies in the way of words*, MA: Harvard University Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). Some, and possibly all, scalar inferences are not delayed: evidence for immediate pragmatic enrichment. *Cognition 116*, 42–55.

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. New York, NY: Wiley-Blackwell.

Hertrich, I., Kirsten, M., Tiemann, S., Beck, S., Wühle, A., Ackermann, H., & Rolke, B. (2015). Context dependent impact of presuppositions on early magnetic brain responses during speech perception. *Brain and Language*, *149,* 1-12.

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

Huang, Y., & Snedeker, J. (2011). Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes*, *26*, 1161-1172.

Jouravlev, O., Stearns, L., Bergen, L., Eddy, M., Gibson, E., & Fedorenko, E. (2016). Processing temporal presuppositions: An event-related potential study. *Language, Cognition and Neuroscience, 31,* 1245–1256.

Kirsten, M., Tiemann, S., C. Seibold, V., Hertrich, I., Beck, S., & Rolke, B. (2014). When the polar bear encounters many polar bears: Event-related potential context effects evoked by uniqueness failure. *Language, Cognition and Neuroscience*, *29*, 1147-1162.

Levinson, S. (2000). *Presumptive meanings*. Cambridge, England: Cambridge University Press.

Magri, G. (2010). *A theory of individual-level predicates based on blind mandatory scalar implicatures*. Dissertation, Massachusetts Institute of Technology.

Noveck, I. (2018). How logical terms can be enriched: Exposing semantic-pragmatic divergences. In *Experimental pragmatics: The making of a cognitive science* (pp.78-101). Cambridge: Cambridge University Press.

Noveck, I. A. & Sperber, D. (2007). The why and how of experimental pragmatics: The case of 'scalar inferences.' In N. Burton-Roberts, N. (Ed.), *Pragmatics*. (pp.184-212). UK: Palgrave Macmillan.

Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: an evoked potentials study. *Brain and Language*, *85*, 203-10.

Pouscoulous, N., Noveck, I. A., Plitzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in scalar implicature calculation. *Language Acquisition*, *14*, 347-375.

Romoli, Jacopo. (2011). The Presuppositions of Soft Triggers aren't Presuppositions. *Semantics and Linguistic Theory, 21*, 236-256.

Sauerland, U. (2004) Scalar implicatures in complex sentences. *Linguistics and Philosophy 27,* 367-391.

Schlenker, P. (2011) DRT with local contexts. *Natural Language Semantics 19*, 373–392.

Schneider, C., Schonard, C., Franke, M., Jäger, G., Janczyk, M. (2019) Pragmatic processing: An investigation of the (anti-)presupposition of determiners using mouse-tracking. *Cognition 193,* 104024.

Schneider, C. Bade, N., Franke, M., & Janczyk, M. (2020). Presuppositions of determiners are immediately used to disambiguate utterance meaning. *Psychological Research*, *44*, 1931–1946.

Schneider, C., Bade, N., & Janczyk, M. (2020). Is immediate processing of presupposition triggers automatic or capacity-limited? *Journal of Psycholinguistic Research. 49,* 247-273.

Schwarz, F. & S. Tiemann (2012). Presupposition processing - the case of german wieder. In M. Aloni, V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz, and M. Westera (Eds.), *Logic, Language and Meaning: 18th Amsterdam Colloquium* (pp. 200-209) Berlin: Springer-Verlag.

Schwarz, F. & Tiemann, S. (2017) Presupposition Projection in Online Processing, *Journal of Semantics*, *34*, 61–106.

Schwarz, F. (2007). Processing presupposed content. *Journal of Semantics 24*, 373-416.

Schwarz, F. (2016). Experimental work in presupposition and presupposition projection. *Annual Review of Linguistics, 2*, 273–292.

Sedivy, J., Tanenhaus, M., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretations through contextual representation. *Cognition, 71,* 109-147.

Simons, M. (2001). On the conversational basis of some presuppositions. In Hastings, R., Jackson, B., Zvolenszky, Z. (Eds.) *Proceedings of Semantics and Linguistics Theory 11,* (pp. 431-448), New York: New York University.

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic, 2,* 447–457.

Strawson, P. (1952*). Introduction to logical theory*. Routledge Revivals. Routledge.

Tiemann, S. (2014). The Processing of wieder (‘again’) and other Presupposition Triggers, Dissertation, University of Tübingen.

Tiemann, S., Kirsten, M., Beck, S., Hertrich, I., & Rolke, B. (2015). Presupposition processing and accommodation: An experiment on wieder (‘again’) and consequences for other triggers. *Studies in Theoretical Psycholinguistics, 45*, 39-65.

Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., Knapp, J., & Beck, S. (2011). Psycholinguistic evidence for presuppositions: On-line and off-line data. In I. Reich, E. Horch, & D. Pauly (Eds.) *Proceedings of Sinn & Bedeutung 15* (pp. 581–595). Saarbrücken: Saarland University Press.

van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*, 235–248.

van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147–182.

von Fintel, K. (2004). Would you believe it? The King of France is back! (Presuppositions and truth-valueintuitions). In M. Reimer & A. Bezuidenhout (Eds.), *Descriptions and beyond* (pp. 315–341). Oxford: Oxford University Press

Appendix

Table 2: Regions of interest for the respective conditions

	P1: **PSP trigger/scalar term**	**P1+1:** **post-trigger**	**P2:** **evaluation**	**P2+1:** **post-evaluation**	**P2+2:** **post-evaluation 2**	**P3:** **final word**
Context	$F(1,47) = 0.71$, $p = .405$, $\eta_p^2 < .01$	**$F(1,47) = 6.91$. $p = .012$, $\eta_p^2 = .13$**	$F(1,47) = 0.00$, $p = .980$, $\eta_p^2 < .01$	$F(1,47) = 4.59$, $p = .037$, $\eta_p^2 = .09$	**$F(1,47) = 9.44$ $p = .004$, $\eta_p^2 = .17$**	$F(1,47) = 3.69$, $p = .061$, $\eta_p^2 = .07$
Sentence type	**$F(3,141) = 79.51$, $p < .001$, $\eta_p^2 = .63$, $\varepsilon = .69$**	**$F(3,141) = 180.70$, $p < .001$, $\eta_p^2 = .79$, $\varepsilon = .58$**	**$F(3,141) = 174.55$, $p < .001$, $\eta_p^2 = .79$, $\varepsilon = .47$**	**$F(3,141) = 30.74$, $p < .001$, $\eta_p^2 = .40$, $\varepsilon = .75$**	**$F(3,141) = 52.00$, $p < .001$, $\eta_p^2 = .53$, $\varepsilon = .80$**	**$F(3,141) = 42.91$, $p < .001$, $\eta_p^2 = .48$,** $\varepsilon = .62$
Interaction	**$F(3, 141) = 4.60$ $p = .012$, $\eta_p^2 = .09$**	**$F(3,141) = 5.35$ $p = .004$, $\eta_p^2 = .10$, $\varepsilon = .76$**	**$F(3,141) = 10.36$, $p < .01$, $\eta_p^2 = .18$, $\varepsilon = .55$**	**$F(3,141) = 9.66$, $p < .001$, $\eta_p^2 = .17$, $\varepsilon = .84$**	**$F(3,141) = 6.69$ $p = .001$, $\eta_p^2 = .12$, $\varepsilon = .75$**	$F(3,141) = 1.93$, $p = .128$, $\eta_p^2 = .04$
determiner	$F(1,47) = 0.32$, $p = .576$, $\eta_p^2 = .01$	**$F(1,47) = 6.99$, $p = .011$, $\eta_p^2 = .13$**		**$F(1,47) = 13.94$, $p < .001$, $\eta_p^2 = .23$**	$F(1,47) = 3.13$, $p = .083$ $\eta_p^2 = .06$	
again	$F(1,47) = 0.34$, $p = .562$, $\eta_p^2 = .01$	$F(1,47) = 0.55$, $p = .464$, $\eta_p^2 = .01$	$F(1,47) = 0.98$, $p = .328$, $\eta_p^2 = .02$	$F(1,47) = 1.21$, $p = .277$, $\eta_p^2 = .03$	$F(1,47) = 2.74$, $p = .105$, $\eta_p^2 = .06$	
SI *some*	**$F(1,47) = 5.58$, $p = .022$, $\eta_p^2 = .11$**	**$F(1,47) = 4.96$, $p = .031$, $\eta_p^2 = .10$**	**$F(1,47) = 17.50$, $p < .001$, $\eta_p^2 = .27$**	**$F(1,47) = 6.77$, $p = .012$, $\eta_p^2 = .13$**	**$F(1,47) = 5.50$, $p = .023$, $\eta_p^2 = .10$**	
SI *all*	**$F(1,47) = 5.55$, $p = .023$, $\eta_p^2 = .11$**	$F(1,47) = 3.69$, $p = .061$, $\eta_p^2 = .07$	**$F(1,47) = 7.79$, $p = .008$, $\eta_p^2 = .14$**	**$F(1,47) = 9.55$, $p = .003$, $\eta_p^2 = .17$**	**$F(1,47) = 13.44$, $p = .001$, $\eta_p^2 = .22$**	

Table 3. Inferential statistics for Experiment 1. The first rows are the statistics for the 4×2 ANOVA for each region. In case of a significant interaction, separate ANOVAs with context condition as a repeated measure were run.

					P 1	P 1 +1		P 2	P 2+1	P 2+2			P 3
determine			Manuel	holt	**das/die**			Ticket/ s	und	freut			sich.
again		Heute	hat	Lukas	**wieder (keine)**	Pizza		bestellt	und	hat	nichts	zu	essen.
SI	***all***			Weil	**einige**	Schrau-ben	kaputt	sind	müssen	neue	gekauft		werden..
	some			Weil	**alle**	Schrau-ben	kaputt	sind	müssen	neue	gekauft		werden..

Note: P1 = PSP trigger/scalar term; P2 = word of evaluation, P3= final word