Presupposition Projection in Online Processing

Florian Schwarz and Sonja Tiemann
Department of Linguistics and IRCS, University of Pennsylvania and
Eberhard Karls Universität Tübingen

May 22, 2015

(Revised version)

Abstract

A central aspect of language comprehension is that hearers integrate incoming linguistic content both with the rest of the current sentence and the larger discourse context. Presuppositions crucially interact with both intra- and inter-sentential context in intricate ways, which makes their study especially useful in this regard. We present a series of experiments investigating the time-course of interpreting presuppositions in online comprehension and the impact that so-called presupposition projection in cases where presuppositions appear in embedded environments has on this. We find immediate delays in eye tracking reading times when the presupposition of German *wieder* ('again') is not supported by the context, but only for unembedded occurrences of *wieder*. Further evidence from a rating experiment and a stops-making-sense study supports our interpretation of this result to the effect that global presuppositions of embedded presupposition triggers are not immediately available in processing. A second reading time experiment explores the effects of embedding further by providing presuppositional support in different locations in contexts with a more complex structure involving conditionals. We find longer reading times when the support is more distant, measured in terms of the number of projection steps posited by Discourse Representation Theory (DRT). Altogether, the results suggest that presupposition projection is a cognitively effortful process, and are thus consistent with theoretical accounts that reflect this in terms of the complexity of the representations involved in the different types of contexts, while other accounts that are more neutral in this regard need to be supplemented by additional assumptions or alternative explanations for the observed effects. On the more general level of discourse processing models, these results suggest that there is even more structure relevant to cognitive processes at a level between the surface representation and the purely semantic level (e.g., the commonly assumed level of a text-base) than previously assumed.

1 Introduction

One important question in the study of natural language meaning is how comprehenders piece together the overall conveyed meaning of an utterance in context. Much work in philosophy of language and linguistic semantics has argued that we need to distinguish theoretically between levels of meaning such as the literal truth-conditions, presuppositions, and conversational implicatures (as well as conventional implicatures, which some authors see as on par with presuppositions). From a theoretical perspective, many interesting questions arise as to what exactly the classes of meanings to be distinguished are, how these various ingredients of the overall conveyed meaning arise, how they interact, and how they behave in different linguistic environments. From the perspective of online processing, equally interesting questions concern the when and how of these aspects of meaning being computed in comprehension. One question, which has already received substantial attention in the recent experimental literature on implicatures, is how the different types of meaning relate to one another in online processing. Another question is how aspects of meaning are related to and integrated with both other information in the same sentence as well as the overall discourse context. The psychological literature on discourse processing includes various proposals for how situation models, seen as language-independent representations of the expressed content, are constructed from linguistic input. One of the central upshots from this literature is that there are cognitively real levels of representation that are neither identical to the linguistic surface form nor as independent from the linguistic structure of the surface form as situation models, as they maintain, say, a structured representation of propositions in the form of predicates and their arguments. Understanding precisely how information based on linguistic input is structured at this intermediate level is central for a cognitive theory of language comprehension in context.

The experiments reported in this paper focus on the online processing of the presupposition introduced by German *wieder* ('again'), by looking at sentences such as the one in (1), where *again* introduces the presupposition that Tina had been ice-skating before (see below for more detailed discussion of the properties of presuppositions).

(1) On Saturday, Tina went ice-skating again.

Presuppositions are of particular interest in light of the general questions raised above. First, while presuppositions are a formally well-studied aspect of meaning, little is known about their online processing, especially in comparison with other aspects of meaning, such as implicatures and asserted content. Even more importantly, presuppositional phenomena exist right at the interface between the intra-sentential computation of meaning (typically seen as the realm of formal semantics in theoretical terms) and the broader integration of sentential meaning into the discourse context (part of a theory of discourse, or pragmatics more generally). The core

phenomenon of interest in this regard is that of presupposition projection, which arises when presuppositional expressions are introduced in embedded contexts and seems to involve - loosely speaking - a mismatch between the syntactic location of the expression introducing a presupposition and the level at which the presupposition makes an impact (more details below). Our experiments investigate the time-course of presupposed content in both embedded and unembedded environments. In the experiments presented in this paper, we focus on the presupposition of *wieder* ('again') exclusively. This choice affords us a minimal manipulation of the order of negation and the presupposition trigger and the corresponding variation in embedding. Moreover, we think it prudent to focus on one trigger at a time since presupposition triggers might very likely vary in their properties and thus in the way they are processed (Simons, 2001; Abusch, 2010; Abrusán, 2011; Domaneschi, Carrea, Penco, & Greco, 2013, among others). Nevertheless, throughout this paper we will often talk about presuppositions in general. For the time being, we take the results obtained for *again* as a starting point and assume that the projection of other presuppositions proceed in a similar fashion. Whether or not these results can be generalized to all other presupposition triggers remains to be determined by future research.

We begin by introducing the basic theoretical and experimental background for our later discussion. We discuss two prominent semantic theories of presupposition - DRT and dynamic semantics - and processing hypotheses that they are naturally associated with. We then discuss two sets of experiments, both of which are concerned with German *wieder* ('again') in unembedded and embedded environments. In the first experiment, we look at eye tracking reading times in contexts that either do or do not support the presupposition of *wieder*, which in turn is introduced either in an unembedded position or in the scope of negation. We find immediate slow-downs in eye tracking reading time measures when the context does not match an unembedded presupposition, indicating that in these cases, the presupposition is assessed relative to information present in the context right away. Intriguingly, such effects are absent when the presupposition trigger is embedded under negation. Two follow up studies, one involving a simple acceptability rating task and one using a 'stops making sense' task, are presented as well to narrow down the interpretation of the reading time results. Taken together, the pattern in the experimental results is such that we see immediate effects in conditions involving unembedded presuppositions, while effects are delayed (or absent) in conditions involving presuppositions embedded under negation.

This interpretation of the first set of studies yields straightforward predictions for more complex embedding environments. The second set of studies introduces a larger range of embeddings, both in terms of the embedding expressions involved and the depth of embedding, and varies the level at which the presupposition is resolved. We find significant increases in reading times based on the hierarchical distance between the location where the presupposition is introduced and where it

is interpreted. Our proposed interpretation of these results is that the evaluation of presupposed content in online processing takes place relative to a structured representation of information introduced in the discourse rather than a purely semantic representation that is unstructured. But we also consider potential alternative interpretations in order to fairly assess the theoretical implications of our experiments. The final section draws some broader connections to issues in discourse processing and presents some general conclusions.

2 Presuppositions: Theoretical and Experimental Background

2.1 The Classic Picture - Basic Properties of Presuppositions

The probably most influential tradition in presupposition theory, following work by Karttunen and Stalnaker (Karttunen, 1973; Stalnaker, 1973, 1974, 1978) sees presupposition triggers as introducing constraints on the contexts in which a given sentence can be uttered felicitously. In particular, they express information that has to be entailed by what counts as established in the discourse. While on some level, this clearly suggests a pragmatic view of presuppositions in that they place restrictions on the use of linguistic expressions, different strains of theories in this tradition vary in their take on this: some see presuppositions as part of what is conventionally encoded in the lexical entries of certain expressions; others assume that presuppositions are derived in an entirely pragmatic fashion instead, based on general reasoning much like that involved in generating conversational implicatures.

Historically speaking, there originally was a trajectory from pragmatic to semantic accounts as precise formal implementations of presuppositional phenomena were developed in linguistics. In particular, the work of Heim (1983) and Kamp (1981), to be discussed in more detail later, introduced a shift that incorporated contextual aspects of interpretation into the semantic realm. However, the recent surge in new variations of presupposition theories has come with a revival of more purely pragmatic accounts (Simons, 2001; Abusch, 2010; Schlenker, 2008b, 2009; Abrusán, 2011).[1]

Independent of the debate about the semantic vs. pragmatic status of presuppositions, the probably most central characteristic of presuppositions is that they commonly escape the scope of certain operators. For example, all of the following variations of the simple sentence in (2) seem to have the same presupposition, introduced by *again*, that Tina had been ice-skating before:[2]

[1]Note that some of these types of accounts, perhaps most prominently Abusch's, assume that there are different types of presupposition triggers, and may not be intended to apply to triggers like *again*. But since this restriction does not hold for all of these types of accounts, it's fair to consider the class as a whole in light of the present experimental investigation.

[2]Note that the exact nature of the presupposition for (b) is controversial, as dynamic semantic

(2) Tina went ice-skating again today.

a. It is just not true that Tina went ice-skating again today.
b. If the weather was nice, Tina went ice-skating again today.
c. Did Tina go ice-skating again today?

Unlike the plain declarative version in (2), the asserted content that Tina went ice-skating today is no longer conveyed by these variations due to the effect of the various embedding operators: negation in (2-a) leads to the assertion of just the opposite; the conditional in (2-b) considers the possibility of the truth of the assertion in particular circumstances characterized by the *if*-clause; and the question in (2-c) requests information as to whether this content holds or not. But the presupposition remains unaffected by all of these embedding operators. This phenomenon of presuppositions contributing to the interpretation of an utterance without being affected by various types of embedding operators is generally known as 'presupposition projection' (Karttunen, 1973). Throughout this paper we will use the term *global interpretation* for the relevant readings. They require the presupposition to be established as true in the overall context, or else presupposition failure will occur. In the most simplistic terms, presupposition failure is a mismatch of what is implicitly assumed in a sentence with what is given in the context. However, at least since Lewis (1979), it has been well-known that presuppositions can sometimes be accommodated in such circumstances, i.e., be added to the context after encountering the sentence.[3]

Matters are even more complicated than illustrated in (2), in that presuppositions can interact with other content in their sentence so that the contextual requirements otherwise present at the global level sometimes disappear. For example, the overall conditional sentence in (3), which has the same consequent as (2-b), does not seem to come with any presupposition requiring that it counts as established in the context prior to its utterance that Tina had been ice-skating before:

(3) If Tina went ice-skating last week, then she went ice-skating again today.

This obviously has to do with the fact that an occasion of Tina ice-skating is mentioned in the *if*-clause here, but formulating a theory that systematically derives these phenomena is by no means trivial. A family of theories originating in the early 1980's proposed to solve this issue by incorporating aspects of contextual interpretation into the semantic representation. These theories see the denotation of

accounts assume it to be conditional.

[3]Note that global accommodation is crucially different from the standard case of a global interpretation, where the global context supports the presupposition. If the context does not support it, the presupposition can either be accommodated (which involves an adjustment of the context resulting in a new context that does support it (Lewis, 1979)) or the whole sentence ends up as infelicitous (Strawson, 1950). For embedded occurrences of presupposition triggers, there may be the additional option of local accommodation, which is standardly seen as a last resort; see discussion below.

sentences as encoding the impact that its utterance would have on any given context. They are typically grouped together as broadly speaking 'dynamic' theories.

We will focus on two classical instances of such theories, namely Dynamic Semantics, which started with Heim (1983) (Groenendijk & Stokhof, 1990; Chierchia, 1995, among many others), and DRT (Kamp, 1981; van der Sandt & Geurts, 1991; van der Sandt, 1992; Kamp & Reyle, 1993; Geurts, 1999). While similar in spirit in that they formally incorporate the discourse context into the process of semantic interpretation, they crucially differ in that DRT adds an additional representational level of so-called 'Discourse Representation Structures' (DRSs), whereas Dynamic Semantics sees the semantic effect of sentences as directly updating the information encoded by the context in a non-representational way. Both theories offer accounts of presupposition projection that overlap substantially in their predictions, and are largely successful in capturing the empirical data. However, given the difference in the representational levels involved, the two accounts are most naturally associated with quite different processing hypothesis, as will be illustrated in section 2.4.

2.2 The Experimental Investigation of Presuppositions in Context

While presuppositional phenomena have played a role in a number of experimental investigations of language processing for quite some time, e.g., in connection with syntactic parsing decisions (Crain & Steedman, 1985) and reference resolution of definites (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995, and a large subsequent literature in the visual world paradigm), they have only become a focus of psycho-linguistic work on meaning more recently. The past few years have seen a growing body of work that tries to assess the interpretive properties of presuppositional content, much of it using off-line behavioral measures. Some of the issues addressed in the literature include the strength of contextual constraints imposed by a variety of presupposition triggers (Jayez & van Tiel 2011, Amaral et al. 2011, Smith & Hall 2011), the effects of presuppositional content on resolving ambiguities (Schwarz, 2007), and the exact nature of presuppositions in conditionals and under quantification (Chemla & Schlenker, 2012; Chemla & Bott, 2013; Romoli, Sudo, & Snedeker, 2011).[4]

There also is a growing body of work trying to understand the online processing of presuppositions at a more detailed level by looking at various measures reflecting the time course of presupposition interpretation in online processing. Based on the general notion that presuppositions require some form of contextual support, previous studies have explored experimental methods for investigating presuppositions. Schwarz (2007) found reading time effects for the part of the sentence containing *also* when the preceding sentential context did not support its presupposition. Building on this paradigm, Tiemann et al. (2011) investigated a broader range of presupposition triggers and found that unsupported presuppositions gave

[4]For a recent collection of contributions in this area, see (Schwarz, 2015a).

rise to decreased acceptability and increased reading times on the presupposition trigger itself in word-by-word self-paced reading. The self-paced reading studies of Schwarz (2007) and Tiemann et al. (2011) found increases in reading times upon seeing the expressions giving rise to the presupposition when the context did not support the presupposition. These increases are attributed to the clash between the context and the presupposition trigger and thus can be seen as indicative of the availability of the presuppositional content, since a mismatch can only be noticed when the presupposition has been fully computed. More recent results from the visual world paradigm furthermore suggest rapid interpretation of the presupposition of *also* (Romoli, Khan, Snedeker, & Sudo, 2015; Schwarz, 2015b), as well as of *again* and *stop* (Schwarz, 2014). Finally, Chemla and Bott (2013) investigate reaction times for various interpretive options of presupposition triggers like *realize* under negation and report that global interpretations are faster than local ones (see section 3.1.4 and section 5.2.4 for discussion of local accommodation).

Theoretically speaking, one approach to the interpretation of findings about the time-course of interpreting presuppositions, in particular in comparison to asserted content, is to take a perspective broadly parallel to the literature on scalar implicature processing. Many authors (Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; Huang & Snedeker, 2009, 2011; Schwarz, Romoli, & Bill, 2014) have argued that the pragmatic strengthening of *some* to *some but not all* (Grice, 1975) is a process that takes place online and comes with a processing delay relative to the literal *some and possibly all* interpretation. From this angle, pragmatic theories of presupposition would seem to suggest that presuppositions, too, are slow and costly to derive in online processing. Semantic presupposition theories, on the other hand, would lead us to expect that presuppositions are processed at least as early as asserted content, if not before.

However, this perspective likely is both too simplistic on conceptual grounds and contentious in empirical terms: the evidence for delayed availability of implicatures is by no means uniform, as various authors have presented evidence for rapid implicature computation (Grodner, Klein, Carbary, & Tanenhaus, 2010; Breheny, Ferguson, & Katsos, 2013); and the assumption that pragmatically generated content is only available with a delay is by no means a necessary one. The latter point is particularly relevant for our finding of immediate reflexes of unembedded presuppositions in processing, which are thus compatible either with a semantic account or a pragmatic one that assumes rapid pragmatic processing. Importantly, however, the parallels between implicatures and presuppositions only go so far, and the projection phenomena for embedded cases add another angle to this debate. In particular, at least certain pragmatic accounts of presupposition projection do not seem to be compatible with the overall pattern of experimental results (see also Chemla & Bott, 2013, for highly relevant results, which are reviewed in section 5.2.4).[5]

[5]More specifically, most pragmatic accounts of recent fall in the same category as Dynamic Semantics in this regard. While they might lead us to expect that presuppositions in general are

2.3 Presuppositions and Projection in Two Classical Theories

Leaving aside potential differences in processing implications between semantic and pragmatic accounts, semantic accounts arguably differ in terms of the processing hypotheses that they most naturally associate with, specifically with regards to presupposition projection. To begin with, consider what is involved in projection on a general level. Descriptively speaking, projection could be characterized as a mismatch between the level at which an expression is introduced syntactically and the level at which it makes an impact. Depending on how this effect is captured in an incremental comprehension system, dealing with this type of mismatch could well involve additional efforts, as the different aspects of interpretation have to be sorted out and integrated at the appropriate level relative to the relevant embedding operators. In the following, we consider the two standard theories of projection, DRT and Dynamic Semantics, and formulate processing hypotheses that are based on the theoretical mechanisms they posit.

Some caveats are in order. First, these are by no means the only theories aiming to account for presupposition projection. In particular, there are various recent pragmatic and trivalent accounts of presuppositions that also deserve consideration here. Some possible interpretations of (some of) the data in line with these theories are considered in section 5.3. We focus on the two present theories because they capture context update dynamically, which lends itself to an extension to sentence processing. Furthermore, they provide well-developed formal frameworks that help to illustrate a key dimension along which processing models of presupposition interpretation can differ. That being said, it is clear that these two theories are not necessarily formulated as processing theories, and they are by no means necessarily tied to the processing hypotheses we associate them with below. Additional factors based on more general and possibly orthogonal processing issues may need to be considered in developing a fully fleshed-out picture of how the processing of presupposition projection actually proceeds, which will likely interact with possible constraints imposed by the theoretical approaches. Nonetheless, we find it fruitful to begin exploring plausible processing hypotheses that build on the theoretical mechanisms inherent in the two accounts. Our discussion of theoretical consequences of our findings will include consideration of possible additional factors that could help to reconcile theoretical options with the results. In the course of this, we will also consider some other theoretical proposals for accounting for projection phenomena.

2.3.1 Discourse Representation Theory (DRT): Projection as operations on representations

The key feature of DRT is that it provides a mechanism for incorporating the meaning of new sentences into a representation of the discourse context that encodes

costly to compute, they do not seem to be able to differentiate between different levels of embedding. Also see the General Discussion for more on this issue.

previously expressed information. It does so by translating individual sentences into discourse representations that can be combined with the representation of the prior context. Discourse representations can be conveniently presented using a box format, where the boxes consist of a 'header' and a main body. The header contains a list of discourse referents or reference markers, which are variables. The main body contains descriptive conditions, i.e., it lists predicates that are said to hold of certain of these variables. Presuppositions are seen as essentially anaphoric expressions in DRT, in that they have to be identified with an antecedent somewhere in the discourse structure (van der Sandt & Geurts, 1991; van der Sandt, 1992). They are initially introduced in the DRS corresponding to the clause they are part of syntactically, but a subsequent procedure manipulates the discourse structure based on a search along a well-defined search path for a suitable antecedent. (5) and (6) provide illustrations of the discourse representations for the sentence in (4-a) and (4-b), respectively.[6] Note that these are schematic English illustrations of the German experimental materials used below. Presupposed material is marked by underlining in the discourse representations.

(4) a. Tina AGAIN NOT went ice-skating today.
b. Tina NOT AGAIN went ice-skating today.

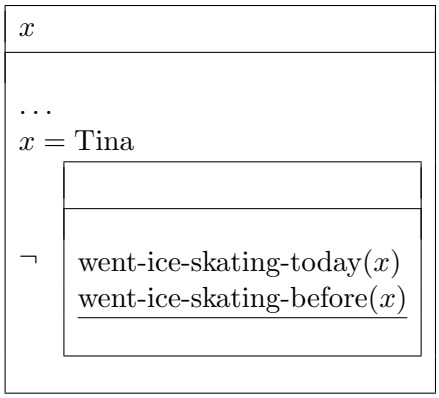
(5)

x
$\ldots$ $x = \text{Tina}$ $\neg\text{went-ice-skating-today}(x)$ $\underline{\neg\text{went-ice-skating-before}(x)}$

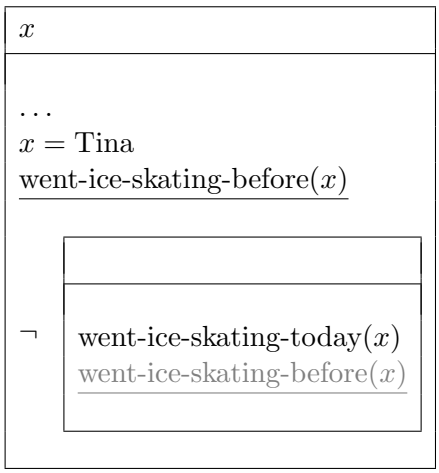
(6)

[6] To avoid the complexities involved in representing events, we simply use complex predicates that can include adverbs like *today*. '...' is used as a place holder for conditions introduced by earlier parts of the discourse. In the case of simple negated conditions, it is sometimes convenient to simply write the condition preceded by negation, without including an extra box.

a.



b.



When *again* is unembedded, its presupposition is introduced at the top level of the discourse representation from the start, i.e., there is no projection. The presupposition then simply has to be resolved by finding a condition introduced in the preceding discourse that matches the presupposition. In other words, it must have been established before the utterance of the sentence that there is some (salient) occasion on which Tina did not go ice-skating. If this is not the case, the presupposition may be accommodated (at least in certain circumstances).[7]

When *again* is embedded under negation (represented by an embedded box in DRT terms), the process is more complicated. Based on its syntactic location, the presupposition of *again* (here that Tina had been ice-skating before) is introduced in the embedded DRS, since that corresponds to the clause it is part of syntactically ((6)a). However, the mechanism responsible for handling projection places the underlined condition in the global DRS, yielding ((6)b). (To keep track of its origin,

[7]This may or may not be an option for *again* - Tiemann (2014) presents experimental evidence that the presupposition of *wieder* is generally not accommodated. In any case, accommodation is not an option in our experimental materials as the discourse context is incompatible with the presupposition in the infelicitous condition.

we will use gray renderings of the condition in its previous location(s).) Now that the presupposition is part of the global DRS, its resolution proceeds in the same fashion as for the unembedded case. It should be noted that the formalism of DRT in general is not necessarily tied to a sequential analysis of these steps. As Geurts (1999, p. 55) notes, it is perfectly possible to assume an under-specified representation that initially leaves open which DRS an embedded presupposition associates with. However, for purposes of formulating a processing hypothesis building on DRT below, we will adopt the original step-by-step projection procedure of van der Sandt (1992) and Geurts (1995).

2.3.2 Dynamic Semantics: Projection via constraints on context updates

The second theory which we will consider in connection with the processing of embedded presuppositions is that of Dynamic Semantics (Heim, 1982, 1983; Groenendijk & Stokhof, 1990; Chierchia, 1995). This theory takes the perspective suggested by Stalnaker (1973) and Karttunen (1973) that presuppositions are conditions on the *context* of use and incorporates these contextual conditions directly into the semantics. It does this by rethinking the meanings of sentences in terms of the impact they have on any given context - their *context change potential*. Formally speaking, the context is a set of possible worlds in which everything that counts as established by the discourse participants holds to be true. The context change potential of a sentence is understood as the impact it has on the context in terms of further restricting the set of worlds to those where the expressed proposition holds, rendering a new context c'. Negation removes worlds in which the negated proposition holds from the context. An *update* of the context c with the unembedded *again* sentence is illustrated in (7-c).[8] The effect of the negated proposition that is asserted (7-a) is to remove those context worlds in which Tina went ice-skating today.

(4-a) Tina AGAIN NOT went ice-skating today.

(7) a. $S' = NOT\ S$, where S: Tina went ice-skating today.
b. PSP of AGAIN NOT S: There's a (salient) earlier time where Tina did not go ice-skating.
c. $c' = c + S' = c + NOT\ S = c - (c + S)$

(defined iff $c + PSP = c$)

The role of presuppositions can now be seen as placing conditions on the definedness of context updates. In particular, an update of c with S' can only take place if its presupposition is already entailed (or *satisfied*) by the context (in which case

[8] The '$-$' symbol here is essentially equivalent to forming the set-theoretic difference, i.e., $c - X$ will yield a set containing all those c-worlds that are not part of X.

adding the presupposed proposition to the context would not remove any worlds). If this is the case, the context *admits* the sentence and can be updated as in (7-a).

One major goal of Dynamic Semantics is to account for the behavior of presuppositions in embedded contexts, e.g. in the scope of negation, conditionals, or quantifiers. Our example target sentence with embedded *wieder* illustrates the case of negation.

(4-b) Tina NOT AGAIN went ice-skating today.

(8) a. $S' = NOT\ S$, where S: Tina went ice-skating today.
b. PSP of AGAIN S: There's a (salient) earlier time where Tina went ice-skating.
c. $c - (c + S)$

(defined iff $c + PSP = c$)

Note that even though the presupposition trigger *again* is part of the negated clause, the definition of the context update for negation ends up requiring the presupposition to hold in the global context c, because it involves the update $c + S$ as its initial step. This is because negation formally involves first updating c with the clause it negates, and then subtracting the result from c. This is precisely what captures the projection effect, in that the presupposition (that Tina has been ice-skating before) is evaluated relative to the original context c, just like it would be if negation were not present.

2.4 Presupposition Projection from a Processing Perspective

We now turn to the question of how these theories might relate to online processing, specifically with respect to the experimental design employed below. The theories we considered are not processing models themselves, of course. However, we believe that they quite naturally associate with certain processing hypotheses, based on their general properties. And given that any comprehensive theory of language comprehension in context will have to tackle the projection problem, some version of a formal account of these phenomena ultimately has to be incorporated into a broader cognitive theory. So theoretical accounts of projection seem like the natural starting point for exploring possible processing models. That being said, it is also clear that it is unlikely that any given theory is only compatible with one particular processing model, or one particular pattern in processing data. Many additional assumptions are necessary about language processing more generally, and additional factors that may or may not be related to the specifics of presupposition projection may come into play in experimental data. We will proceed by spelling out two plausible processing hypotheses that broadly align with certain formal properties of the theoretical accounts under consideration. These hypotheses will then be assessed in our experiments. We return to a discussion of theoretical implications

in the general discussion.

Starting with a hypothesis based on DRT, it is worth noting that the structural representations of this theory are commonly intended to correspond to (or at least model parts of) cognitively real representations of the discourse. To the extent that we assume the process of presupposition projection that we illustrated in (6) to correspond to actual cognitive processes involved in deriving the appropriate interpretations for embedded occurrences of presupposition triggers, this could plausibly be expected to involve additional processing effort and time, compared to unembedded occurrences.[9] Thus we can formulate the following processing hypothesis based on the representational complexity involved in a DRT account:

(9) **Projection-Takes-Time Hypothesis (PTT)**
Processing the presupposition of an embedded trigger whose presupposition projects involves additional steps in interpretation, compared to unembedded ones, which give rise to additional processing cost, as reflected in increased processing time. (This processing cost should correlate with the number of steps involved; see section 4.)

A straight-forward processing implementation of Dynamic Semantics, on the other hand, does not lead to added complexity due to 'projection' in general, insofar as it simply evaluates presuppositions relative to their local contexts, which typically include (or in the case of negation, consist of) the global context. For example, it is part and parcel of the definition of negation in dynamic semantics that we evaluate the presuppositions of a negated sentence relative to the global preceding context c. The process of evaluating the presupposition in the two cases is thus exactly the same, as illustrated in (4-a) and (4-b), and consists of checking whether c entails the presupposition of *again*. Given this equivalence, no difference in processing complexity concerning the evaluation of the respective presuppositions is expected.[10] Thus, based on the properties of the formal account alone, we can consider an alternative processing hypothesis:

(10) **Immediate-Global-Impact (IGI)**
Processing presuppositions consists of checking that they are entailed by their local contexts. Evaluation relative to equivalent local contexts involves equivalent processing efforts. To the extent that the global context features as part of the local context, the global context has an immediate impact on presupposition evaluation, and presupposition evaluation in corresponding

[9] This is based on the step-by-step projection as outlined above. The alternative underspecification account (see section 2.3.1) might suggest slightly different effects based on embedding, perhaps without any impact of the distance to the resolution site.

[10] The only difference is that the presupposition in (4-a) is itself 'negated' (loosely speaking; it's just a proposition at the semantic level, of course). But if anything, that might lead us to expect that case to be harder, if there were any extra efforts involved in dealing with negation (e.g., because the update procedure is more complex).

embeddings does not incur additional processing costs of its own.[11]

This leaves open the existence of independent factors affecting processing that could be held responsible for any observed delays or differences more generally. Furthermore, there may be alternative processing implementations that make slightly different predictions (see section 5.2.2 for discussion). Our investigation of which of these two hypotheses is correct therefore by no means provides a direct evaluation of the theories that these hypotheses were based on. We consider possible additional factors as well as the broader theoretical picture in the General Discussion section.

To sum up, the two semantic theories considered here, despite being largely on par in terms of capturing the central facts about presupposition projection, suggest different processing hypotheses about the costliness of projection in processing. The representational structure posited by DRT and the increased complexity introduced by the projection mechanism are the key factor behind these differences: if the DRT mechanism corresponds to something cognitively real, we expect global interpretations of embedded presuppositions to be more costly than unembedded presuppositions. Dynamic Semantics, on the other hand, does not utilize structured levels of meaning representation where the global context could fail to make an immediate impact (at least if it's part of the relevant local context). It's formal properties therefore do not provide any grounds for projection requiring additional effort. Our experiments aim to assess the two corresponding processing hypotheses as laid out above.

A final note regarding the issue of the general time course of presupposition evaluation in unembedded cases: both of the accounts (at least in their standard form) are compatible with the notion that presupposed content is available at least as early as literal, truth-conditional content. Dynamic Semantics could perhaps be seen to make an even stronger prediction in this regard, namely that presuppositions are evaluated *before* literal, truth-conditional content, since presuppositions have to be checked prior to updating the context with the asserted information (Beaver, 2001). Our experiments will not speak directly to this last issue.

3 *Wieder* - Embedded and Unembedded

The experiments presented here follow the general approach of the reading time studies discussed earlier in that they involve target sentences containing a presupposition trigger, German *wieder* ('again'), presented in contexts that either do or do not support the presupposition. However, we include the use of eye tracking during reading in order to allow for a more natural reading experience for the participants and to have a more fine grained temporal resolution, which allows us to capture effects closer to the real time course. In addition to considering presupposition

[11] Of course the general complexity of embedding may come with its own processing costs independent of presupposition evaluation.

triggers in simple sentences without any embedding, we also consider embedding environments, in particular negation. (Experiment 2 in section 4 adds embedding in conditionals). This allows us to investigate time course effects related to presupposition projection.

3.1 Experiment 1a: Costs of Presupposition Failure in Reading

3.1.1 Design & Materials

The first experiment manipulated whether or not the presupposition trigger *wieder* ('again') is embedded under negation. In implementing this, we took advantage of the syntax of German, where *wieder* ('again') and *nicht* ('not') can appear in adjacent positions in either order. This allowed us to construct target sentences which are minimally different with respect to whether *again* appears inside or outside the scope of negation. We presented such sentences in two different contexts, each of which supported the presupposition of one of the orders of *wieder* and *nicht* while being inconsistent with the other. In the sample item from our materials below, the context sentence (11-a) supports the presupposition of (12-a) (that Tina went ice-skating before), while (11-b) is inconsistent with it. Conversely, (11-b) supports the presupposition of (12-b) (that there was a preceding occasion where Tina did not go ice-skating), while (11-a) is inconsistent with it.[12]

(11) **Contexts**

a. Tina **went** ice skating for the first time with Karl last week. The weather was **beautiful**, and they **had a great time**.

b. Tina **wanted to go** ice skating for the first time with Karl last week. But the weather was **miserable** and they **gave up on their plan**.

(12) **Target Sentences**

a. *Dieses Wochenende war Tina* ***nicht wieder*** *Schlittschuhlaufen,*
This weekend, was Tina not again ice skating
weil das Wetter so schlecht war.
because the weather so bad was
'This weekend, Tina (once) again didn't go ice-skating because the weather was so bad.' [13]
Presupposition: Tina had been ice-skating before.
Assertion: Tina did not go ice-skating this weekend.

b. *Dieses Wochenende war Tina* ***wieder nicht*** *Schlittschuhlaufen,*
This weekend, was Tina again not ice skating

[12] At least on the global interpretation. This generally is taken to be the default, which is supported by our data from Experiment 1b below. See discussion of local accommodation below.

[13] The English paraphrase indicates the intended interpretation with respect to the scope of *again* and negation. The German sentence is unambiguous.

Sentence	Context	Firstword	Felicitous
(12-a)	(11-a)	*nicht*	felicitous
(12-a)	(11-b)	*nicht*	infelicitous
(12-b)	(11-a)	*wieder*	infelicitous
(12-b)	(11-b)	*wieder*	felicitous

Table 1: Overview of Conditions and Factors (Expt1a)

weil das Wetter so schlecht war.
because the weather so bad was
'This weekend, Tina again did not go ice-skating because the weather was so bad.'
Presupposition: There's a salient previous time when she did not go ice-skating.
Assertion: Tina did not go ice-skating this weekend.

The pairing of sentences and contexts yielded a fully counterbalanced 2×2 interaction design with two factors: **Firstword** (whether *wieder* or *nicht* appeared first) and **Felicity** (whether the context supports the presupposition or not). The conditions with the combinations of the levels of the two factors are summarized in Table 1.

3.1.2 Procedure & Participants.

24 sentences with versions for each of the four conditions were created. In addition to the experimental items, there were 48 unrelated filler items. Participants read the sentences on a computer screen while their eyes were being tracked by an EyeLink 1000 eye tracker from SR Research. For half of the items (of both the fillers and experimental sets), participants had to answer yes/no questions, which followed directly after the sentence, to ensure full comprehension of the materials. 32 native speakers of German from the University of Tübingen community participated in the experiment. Participants were split into 4 groups, where each participant saw 6 of the sentences per condition.

3.1.3 Results

The primary focus in our analysis were the reading times on the verb following the {*wieder nicht*} sequence.[14] Since the presupposition of *wieder* crucially relies on the

[14]We use this set-notation to talk about the sequence of *nicht* and *wieder* in either order.

verb of its clause, it only becomes recoverable from explicitly given materials at the point of the verb. Reading times were also examined for the {*wieder nicht*} sequence itself. Standard reading measures were calculated for purposes of analysis. Based on prior self-paced reading experiments using the same general approach (Schwarz, 2007; Tiemann et al., 2011), we expect increases in reading time when sentences are presented in contexts that are inconsistent with the presupposition. The time point at which such increases arise is indicative of the relevant presupposition having been computed at this point.

All analyses used mixed-effect models with participants and items as random effects, using the *lmer* function of the *lme4* package in *R* (Bates, 2005). Given recent arguments by Barr, Levy, Scheepers, and Tily (2013) that maximal random effect structures should be used whenever possible, we generally computed models with the maximal random effect structure that would converge, with random effect slopes for each factor, as well as the interaction where applicable. To assess whether inclusion of a given factor significantly improved the fit of the overall model, likelihood-ratio tests were performed that compared two minimally different models, one with the fixed effects factor in question and one without, while keeping the random effects structure identical (Barr et al., 2013). We report estimates, standard errors, and t-values for all models, as well as the χ^2 and p-value from the likelihood-ratio test for individual factors. To facilitate presentation of results, we will adopt the following shorthand indications for which random effect structure (RES) was used in the *lmer*-syntax in a given case:

- RES-1: $(1 + factor1 * factor2|\ participant) + (1 + factor1 * factor2|\ item)$ (Full model)
- RES-2: $(1 + factor1 * factor2|\ participant) + (1 + factor1 + factor2|\ item)$
- RES-3: $(1 + factor1 + factor2|\ participant) + (1 + factor1 + factor2|\ item)$
- RES-4: $(1 + factor1 + factor2|\ participant) + (1 + factor1/2|\ item)$

For the overall interaction analyses, predictors were centered, so as to render estimates of main effects. Planned comparisons between individual conditions were conducted using the appropriate treatment-coding.

While the comprehension questions did not play a crucial role for the main purpose of the experiment, they serve as an indicator of the extent to which participants did read the experimental materials for full comprehension. Overall accuracy in question answering was 89%, with no significant differences between conditions.

For analyzing reading times, the following standard reading time measures were computed (Rayner, 1998): first fixation duration, which measures the length of the very first fixation on the region of interest (here the verb); go-past time, which here is taken to measure the sum of all fixations on the region of interest prior to any fixations to the right of this region (but not including the time of regressive fixations);

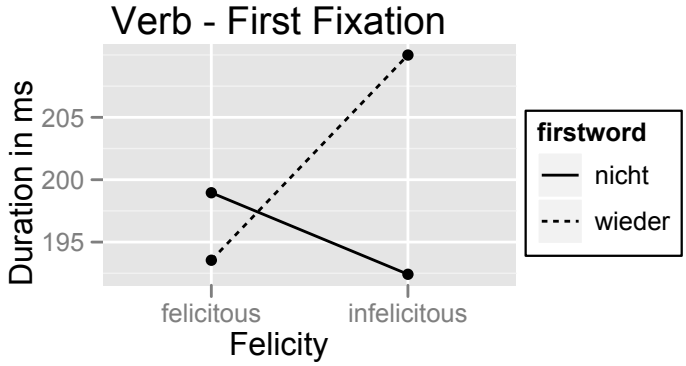


Figure 1: First Fixation Duration (in ms) by condition

first pass time, which includes all fixations on the region when it is looked at the first time, up until leaving the region (to either the left or right); total duration, which sums all the fixations on the region of interest, no matter when they occur; regression path duration, which measures all fixations from first entering the region to first leaving it to the right (including all potential regressive fixations; this is sometimes also referred to as go past time). In addition to the timing measures, we also computed first pass regression proportion, which is the proportion of regressive eye movements following the first time of entering the region. Prior to computing the reading time measures, we removed all trials where a participant blinked at any point while looking at the {*wieder nicht*} + *Verb* region. This resulted in the exclusion of 34 trials, which amounted to 4.4% of the data. The distribution of excluded trials across conditions was roughly even.

Means for the reading time measures on the verb are presented in Table 2. The primary result is an interaction between **Firstword** and **Felicity**, as illustrated by the graph for first fixation durations in Figure 1: when *wieder* was first (i.e., not embedded under negation), reading times on the verb were significantly higher in the infelicitous condition. When *nicht* was first, on the other hand (resulting in *wieder* being embedded under negation), there was no such slow-down (and except in total reading time and first pass time, no significant difference between the felicitous and infelicitous context conditions).

The interaction factor contributed significantly to model fit for first fixation duration (RES-2: $\beta = 23.2$, $SE = 10.4$, $t = 2.23$; $\chi^2 = 4.8$, $p < .05$), go-past (RES-2: $\beta = 117.1$, $SE = 54.7$, $t = 2.14$; $\chi^2 = 4.3$, $p < .05$), total reading times (RES-2: $\beta = 157.6$, $SE = 55.2$, $t = 2.85$; $\chi^2 = 7.4$, $p < .01$), and regression path duration

	wieder nicht		*nicht wieder*	
Reading Measure	felicitous	infelicitous	felicitous	infelicitous
First Fixation	194	210	199	192
Go-Past	292	359	324	285
First Pass	270	281	275	247
Total	309	405	370	307
Reg. Duration	395	619	438	479
Reg. Proportion	17.0%	33.5%	17.4%	20.6%

Table 2: Reading time measures (in ms) and First Pass Regression Proportion (in %) on the verb

(RES-3: $\beta = 212.2$, $SE = 104.6$, $t = 2.03$; $\chi^2 = 4.09$, $p < .05$), and did so marginally for first pass duration (RES-1: $\beta = 41.8$, $SE = 23.5$, $t = 1.78$; $\chi^2 = 3$, $p < .1$). For first pass regression proportion, the interaction did not contribute significantly when random slopes for the interaction were included, but did so marginally when only random intercepts for participants and items were included (RES-3: $z = -1.79$, $p < .1$).[15] The analysis of first pass regression proportions also revealed a main effect of **Firstword** with lower regression proportions in the *nicht wieder* conditions (RES-3: $z = -2.33$, $p < .05$). In addition, there was a main effect of **Felicity** for first pass regression proportion (RES-3: $z = -2.67$, $p < .01$) and regression duration (RES-3: $\beta = 144.9$, $SE = 69.9$, $t = 2.07$; $\chi^2 = 4.08$, $p < .05$). No other main effects contributed to model fit significantly for any of the measures.

The interaction was primarily driven by a simple effect of **Felicity** for the *wieder nicht* conditions, with increases in reading measures for the infelicitous context. This contributed significantly to model fit for first fixation duration (RES-2: $\beta = 17.9$, $SE = 7.5$, $t = 2.4$; $\chi^2 = 5.73$, $p < .05$), regression path duration (RES-3: $\beta = 253.7$, $SE = 89.5$, $t = 2.84$; $\chi^2 = 5.89$, $p < .05$), total time (RES-2: $\beta = 93.7$, $SE = 39.4$, $t = 2.38$; $\chi^2 = 5.25$, $p < .05$), and first pass regression proportion (RES-2: $z = 3.11$, $p < .01$), and did so marginally for first pass time (RES-3: $\beta = 27.6$, $SE = 15.8$, $t = 1.75$; $\chi^2 = 3.03$, $p < .1$) and go-past time (RES-1: $\beta = 37.0$, $SE = 30.1$, $t = 1.23$; $\chi^2 = 3.22$, $p < .1$).

For the *nicht wieder* conditions, the only simple effect of **Felicity** appeared in the total reading time (RES-3: $\beta = 63.7$, $SE = 30.1$, $t = 2.11$; $\chi^2 = 4.41$, $p < .05$), and it was in the opposite direction, with a decreased reading time in the infelicitous condition. Regarding simple effects of **Firstword**, the only effect for the felicitous conditions was for total reading time (RES-3: $\beta = 63.7$, $SE = 30.1$, $t = 2.11$;

[15] For binomial dependent variables, p-values from the *lmer* output are reported here and below.

$\chi^2 = 4.41$, $p < .05$), where reading times were faster in the *wieder nicht* condition than in the *nicht wieder* condition. In the infelicitous conditions, the *nicht wieder* condition displayed faster reading times on the verb than the *wieder nicht* condition for first fixation duration (RES-3: $\beta = 17.5$, $SE = 7.8$, $t = 2.24$; $\chi^2 = 4.85$, $p < .05$), go-past time (RES-2: $\beta = 81.8$, $SE = 37.0$, $t = 2.21$; $\chi^2 = 4.62$, $p < .05$), total time (RES-3: $\beta = 91.6$, $SE = 43.2$, $t = 3.23$; $\chi^2 = 10.32$, $p < .01$) and first pass regression proportion (RES-1: $z = 3.01$, $p < .01$), as well as marginally for first pass time (RES-2: $\beta = 33.6$, $SE = 17.6$, $t = 1.91$; $\chi^2 = 3.49$, $p < .1$).

With respect to reading times on {*wieder nicht*} (taken as a unit, in either order), a parallel interaction effect showed up in the total reading times, with corresponding simple effects of **Felicity** for the *wieder nicht* condition and of **Firstword** for the infelicitous condition. There were no other significant effects in this region.

Given the lack of an effect of **Felicity** in the *nicht wieder* condition on the verb, follow-up analyses on later and larger regions were carried out. No increases in reading times were found for regions consisting of the verb plus 2 following words, the 3 words following the verb, the entire section of the sentence from *nicht wieder* to the end, or, for that matter, for the entire trial duration (i.e., total reading time for the entire paragraph).

3.1.4 Discussion

There are two main points to discuss with respect to the experimental results. The interaction shows that the effect of encountering a presupposition in a context that is inconsistent with it differs based on whether we are dealing with an embedded or an unembedded trigger. Furthermore, the presupposition of unembedded *wieder* gives rise to fairly immediate effects of inconsistency that are reflected throughout a variety of reading time measures. Of particular interest with regards to the latter point are the simple effects for first fixation duration and first pass regression proportion. Already during the first fixation of the verb (which lasts less than 200 ms), the beginning of which arguably is the logically earliest point possible to fully compute the presupposition of *wieder* based on what has been explicitly provided, a 16 ms effect emerges. Based on the experimental design, the increase in the infelicitous condition can be attributed to the inconsistency between the expressed presupposition and the provided context. But for such an inconsistency to arise, the relevant presupposition must of course have been computed. Similarly, the increase in first pass regression proportions indicates that upon first looking at the verb, there is an increased likelihood of returning to look at the preceding context, which is presumably triggered by noticing the same inconsistency. The experiment thus provides evidence that the presupposition of *wieder* is computed rapidly online. As discussed above, such a result could be seen as theoretically compatible either with a semantic view of presuppositions or with a pragmatic view that assumes rapid

pragmatic processing.[16]

Returning to the first point, the picture is rather different for cases where *wieder* is embedded under negation. Assuming the standard global interpretation, the two contexts vary in precisely the same way as was the case for the unembedded occurrence of *wieder*, with one context supporting the presupposition, while the other is inconsistent with it. If the global interpretation of the presupposition were available while reading the verb, we would expect to see an effect on reading times similar to the unembedded condition. However, on none of the reading measures was there a significant increase for the infelicitous condition. In fact, the only significant simple effect (for total reading times) went in the opposite direction. The lack of such an increase thus can be taken as an indication that the global interpretation is not available while the verb is being read.[17]

In principle, there are two possible explanations for why this might be the case. The first would speak directly to deciding between our two processing hypotheses above. In particular, in line with the PTT Hypothesis, deriving a global interpretation in the context of an embedding operator like negation might involve increased processing effort, leading to a delay in the availability of a global interpretation. If this interpretation can be maintained, it would directly speak against the IGI Hypothesis.

There is an alternative interpretation, however, which would reconcile the results with the IGI Hypothesis. It is based on the possibility (glossed over in our discussion so far) of so-called *local* accommodation of presuppositions in the scope of negation. Perhaps the most well-known case of this involves the existence presupposition of the definite article, as in (13). While definite descriptions such as *the King of France* presuppose that there is an individual with the relevant property, such a global interpretation of the presupposition is inconsistent with the continuation in this case. Nonetheless, there is a consistent interpretation of the entire text, which shows that it is possible to interpret the presupposition locally, so that the existence of a King of France is part of what is denied by the negation.

(13) The King of France is not bald - because there is no King of France!

Similarly, in (14) it seems possible to negate the presupposition that Tina had

[16]The extent to which these results generalize to other presupposition triggers remains to be explored. Triggers very well could vary in their behavior, based on proposed theoretical distinctions in the literature (Simons, 2001). But see initial evidence for uniform behavior of *again* and *stop* reported in Schwarz (2014).

[17]Note that the pattern here cannot be attributed to differences in ease or speed of accommodation, as accommodation is of no help in any of the conditions: in the felicitous conditions, the (global) presupposition is supported by the context; in the infelicitous ones, it is inconsistent with it, leading to a clash. Even if local accommodation were involved in the *nicht wieder*-condition - see discussion below -, we would expect differences in outcome for the felicitous vs. infelicitous conditions if we assume differences in ease of accommodation (presumably with delays for local interpretations, based on the results by Chemla & Bott, 2013; Romoli & Schwarz, 2015).

been ice-skating before, rather than the asserted content (note that these interpretations may require a special intonation).

(14) Tina didn't go ice-skating again last weekend - this was the first time!

A simple way of modeling this interpretation is to simply assume that both the presupposed and the asserted content remain in the scope of negation, so that the overall interpretation of the sentence can be paraphrased as follows:

(15) NOT [Tina went ice-skating before AND went ice-skating this weekend]

While in principle, the falsity of either conjunct in the scope of negation would suffice to make this true, the fact that one could express the negation of the second conjunct more straightforwardly (by simply leaving out the presupposition trigger altogether) seems to bias this towards an interpretation where it is indeed the falsity of the conjunct contributed by the presupposition trigger that is conveyed by an utterance of this sentence. Thus, local accommodation gives rise to the claim that Tina had not been ice-skating before.

Given a paraphrase along the lines of (15), the existence of local accommodation for the target sentence in the experimental materials would make the *nicht wieder* sentences perfectly consistent with either context. If the context states that Tina had been ice-skating some time recently, then the global presupposition of course remains consistent with that (note that the local accommodation interpretation is not strictly speaking inconsistent with this either, if the paraphrase above is correct). And if the context states that she did not go ice-skating (and had never done it before, either), then the local accommodation interpretation is perfectly consistent with that.

As we are looking at online processing measures, we also have to consider the time-course of such an interpretation becoming available. If a local accommodation interpretation is immediately available for *wieder* in the scope of negation, then the results indeed may be consistent with the IGI Hypothesis. The immediate presence of an interpretation that is consistent with the context would suffice to account for the absence of a reading time delay, independent of the timing of the availability of a global interpretation.[18]

As a first step towards investigating a possible role of local accommodation, we conducted an acceptability rating study to assess whether a local accommodation interpretation is a viable option for our materials.

[18]Note that existing experimental results on local accommodation (e.g., Chemla & Bott, 2013; Romoli & Schwarz, 2015) suggest that it comes with a significant delay based on response time data (see also section 5.2.4). A delay in availability of both global and local interpretations might be consistent with the data as well, but that option would not help to reconcile the IGI Hypothesis with the data.

3.2 Experiment 1b: Acceptability Rating Study of Presupposition Failure in Embedded and Unembedded Contexts

3.2.1 Design & Materials

If local accommodation is indeed a viable option for the presupposition of *wieder* when it appears in the scope of negation, we would expect this to affect speakers' acceptability judgments of these sentences in the two contexts, since infelicity should have a direct impact on acceptability.[19] In particular, the type of interaction that we saw in the reading times should also be present in the judgments. The infelicitous *wieder nicht* condition should be unacceptable, whereas what we dubbed the 'infelicitous' *nicht wieder* condition should be perfectly acceptable based on a local accommodation interpretation. If local accommodation is not a viable option, on the other hand, the *nicht wieder* sentences in the infelicitous context condition should be judged to be less acceptable than in the felicitous context condition.

3.2.2 Procedure & Participants

A rating questionnaire was conducted via the web using the WebExp2 software (`http://www.webexp.info`). The materials were exactly the same as those used in the eye tracking experiment, including all the fillers. Participants were asked to rate the appropriateness of a given discourse on a scale from 1 (least appropriate) to 5 (most appropriate). Since the only variation for a given context consisted of the content of the presupposition, based on the order of *nicht* and *wieder*, any differences we find in acceptability can be attributed to the status of the presupposition, as before. Data from 24 participants was collected.

3.2.3 Results

The results in form of the mean ratings by condition are summarized in Table 3. While there was a marginally significant interaction between **Firstword** and **Felicity** (RES-3: $\beta = .42$, $SE = .22$, $t = 1.90$; $\chi^2 = 3.57$, $p < .1$), as well as a main effect of **Firstword** (RES-3: $\beta = .52$, $SE = .14$, $t = 3.74$; $\chi^2 = 11.5$, $p < .001$), more importantly for our purposes there was a clearly significant main effect of of **Felicity** (RES-3: $\beta = 1.13$, $SE = .13$, $t = 8.94$; $\chi^2 = 35.71$, $p < .001$), with items containing felicitous contexts getting higher (= better) ratings than those containing infelicitous contexts. While this effect was slightly more pronounced in the *wieder nicht* items (as reflected in the marginal interaction), there nonetheless is a significant simple effect for *nicht wieder* in the same direction (RES-3: $\beta = .92$,

[19] Acceptability is clearly a gradient notion that is affected by a host of factors, including non-linguistic ones. Whether felicity is categorical or gradient (and whether it is distinct from acceptability in the first place) is not a straightforward matter. For present purposes, all that we need to assume, however, is that felicity affects acceptability, with a decrease in felicity leading to a decrease in acceptability.

	wieder nicht		*nicht wieder*	
	felicitous	infelicitous	felicitous	infelicitous
Mean Rating	3.94	2.63	3.23	2.34

Table 3: Results of the rating experiment

$SE = .17$, $t = 5.51$; $\chi^2 = 24.18$, $p < .001$), just as there is for *wieder nicht* (RES-3: $\beta = 1.34$, $SE = .17$, $t = 7.98$; $\chi^2 = 39.2$, $p < .001$). With regards to simple effects of **Firstword**, the *wieder nicht* order was judged significantly better in the felicitous condition (RES-1: $\beta = .73$, $SE = .18$, $t = 4.12$; $\chi^2 = 13.10$, $p < .001$), and marginally better in the infelicitous condition (RES-3: $\beta = .32$, $SE = .18$, $t = 1.76$; $\chi^2 = 3.06$, $p < .1$).[20]

3.2.4 Discussion

The rating results show that for both embedded and unembedded *wieder*, the **Felicity** manipulation had a clear effect and resulted in decreased acceptability when the context sentence was inconsistent with the (global) presupposition of *wieder*. This would be unexpected if local accommodation of *wieder* under negation was a generally viable option for basing one's judgment on. This at least provides partial evidence against an interpretation of the data from Experiment 1a that is consistent with the IGI Hypothesis, based on the availability of local accommodation.

However, this need not mean that a local interpretation is non-existent altogether, as it could just be strongly dispreferred. In the theoretical literature, it is commonly assumed that global interpretations of presuppositions constitute the default option, and thus are preferred over local ones. Note that the dispreference for the local interpretation would have to be so strong that it leads people to judge a discourse unacceptable, even though it offers a possible interpretation that would make it consistent with the context. Since speakers seem to commonly strive towards finding a felicitous and appropriate construal of utterances, as witnessed by the intuitive acceptability of (13) and (14),[21] the dispreference for the local interpretation would have to be stronger than that pressure. But as far as the data from Experiment 1b is concerned, the notion of a preference for global interpretations is entirely consistent with them.

Whether or not local interpretations are in principle available, what is of primary concern for our endeavor is the time-course of the global interpretation becoming

[20] Ratings for filler items ranged from 3.2 to 4.1.

[21] In a similar vein, the experimental results from Abrusán and Szendrői (2013) on definite descriptions under negation could be taken to suggest that local accommodation interpretations are utilized when this makes it possible to judge a sentence as true.

available. If it were used immediately to relate the current sentence to its context, then the lack of an infelicity effect in Experiment 1a remains unexplained. The worry that Experiment 1b tried to address was that perhaps subjects don't interpret the presupposition globally to begin with, but it found that people's acceptability judgments are clearly based on a global interpretation. The next question is whether we can find any positive evidence for the idea that global interpretations are only available with a delay (i.e., the PTT Hypothesis), even though they are the default. While the data from Experiment 1a are consistent with the PTT Hypothesis, they do not provide any direct support for it, since we did not find a delayed slow-down due to infelicity. We thus need further data to illuminate the time-course of the global interpretation of the sentences in question more directly. The rating experiment actually provides a potential first hint in this regard, as there is a numerical interaction in the reaction times for the different conditions, with responses being slower in the *nicht wieder* infelicitous condition than in all others. However, this pattern did not reach significance, which is likely due to the course-grained nature of the response time collection method and the various uncertainties of web-based data collection. The experiment reported in the following section was designed to shed light on the time-course of the availability of global interpretations.

3.3 Experiment 1c: The Time-Course of Infelicity-Judgments

3.3.1 Design & Materials

In the reading time study, we have no direct evidence on what interpretation participants are pursuing at any given point. In the rating study, we do not have any detailed time-course information. The next experiment attempts to tie the two dimensions together by employing a stops-making-sense task, where participants read the same materials in a self-paced fashion, with a word-by-word presentation of the target sentence. In addition to the self-paced reading task, participants were instructed to abort trials by pushing another button as soon as they felt that the text no longer made sense.

The stops-making-sense design allows us to collect information on two dependent variables. One is the rate of stops-making-sense judgments in the conditions where the presuppositions and the context are inconsistent, which further supplements the results from the rating experiment. Secondly, we get fine-grained response-time data that is different from the reading time data in the first experiment in interesting ways: given the presence of an ultimate judgment on the felicity of the stimulus at hand, we are able to assess the time course of the rejection decisions more directly. For cases where participants continue reading on a given region, we get self-paced reading data for the target sentence on a word-by-word basis.

The materials used in this experiment were the same as in the previous two experiments, including fillers.

condition / word	WN1	WN2	Verb	Verb+1	Verb+2	Verb+3
wieder nicht	2	94	85	12	2	4
nicht wieder	9	70	83	17	9	2

Table 4: Number of aborted trials per word

3.3.2 Procedure & Participants

Participants were seated in front of a computer, and were instructed to read through sentences on the screen, which would be presented bit by bit, with advances to the next part being triggered by pressing a key on the keyboard. The sentences were presented in random order using the Linger software (`http://tedlab.mit.edu/ dr/Linger/`). Contexts were presented sentence by sentence, and the target sentence was presented word by word. Participants were instructed that if the text on the screen stopped making sense, they should press another button. Data from 48 participants was collected, all of whom were members of the University of Tübingen community.

3.3.3 Results

In this experiment, our main focus of analysis are the two infelicitous conditions. As far as the frequency of stops-making-sense judgments is concerned, the *nicht wieder* and *wieder nicht* conditions are essentially on par at 79.1% and 82.9% respectively, without any significant difference between them. The felicitous conditions were judged to not make sense about 20% of the time (the relatively high number of false alarms presumably is due to the nature of the task; comparable rates were found for filler items for which there is no obvious issue with respect to whether they make sense). This supplements the results from the rating study in that the infelicity is clearly detectable in the materials, and it seems to be noticed by participants at the same rate for the embedded and unembedded *wieder* conditions. Thus, there is further evidence against the general viability of local accommodation of the presupposition in the former.

Turning to the response time-course in the infelicitous conditions, there are a number of ways of looking at the data. To begin with, the distribution across words during which the trials were aborted was comparable in the *wieder nicht* and *nicht wieder* conditions, with perhaps a slight shift towards the right in the latter, as can be seen in Table 4. In order to evaluate the time-course of stops-making-sense judgments in more detail, what we are after is the time span lasting from initial access to the presupposition to the stops-making-sense button press. Given that more than

condition	from WN2
wieder nicht	2258
nicht wieder	2799

Table 5: Rejection Times (in ms) from second word in *wieder nicht* sequence in aborted trials in infelicitous contexts

a third of the overall stops-making-sense judgments are already made while looking at the second word in the {*nicht wieder*} sequence (presumably because the verb is relatively predictable based on the stimuli) we decided to start the clock with the display of that second word (*nicht* in the *wieder nicht* condition and *wieder* in the *nicht wieder* condition), and then sum all following per-word response times up to the point of the stops-making-sense judgment.[22] The mean rejection times calculated in this way for all trials ending in a stops-making-sense judgment are presented in Table 5. (For the purposes of aggregation and analysis, outliers that were more than 3 standard deviations from their condition mean were removed.) Rejections were significantly slower in the *nicht wieder* condition (2799 ms) than in the *wieder nicht* condition (2258 ms (RES-1: $\beta = 538.0$, $SE = 202.6$, $t = 2.66$; $\chi^2 = 6.49$, $p < .05$).

In light of potential concerns about the fact that the reading times on *wieder* itself are not included in the calculation of the rejection times based on the method just laid out for the *wieder nicht* condition, we also considered another perspective. In particular, we computed rejection times by summing all reading times starting with *wieder* in both conditions. This, however, results in an imbalance, since there is an additional word that has to be read in the *wieder nicht* condition (namely *nicht*). In order to take this into account, we factored out the reading time for the word *nicht* in the most conservative way, namely by calculating average reading times for *nicht* in the *nicht wieder* condition in the felicitous context for each participant. These mean *nicht* reading times were subtracted from the rejection time in each aborted *wieder nicht* trial in the infelicitous context, on a by-participant basis. The resulting rejection times yielded an even greater difference between the conditions, which also was significant (RES-1: $\beta = 641.4$, $SE = 186.2$, $t = 3.445$; $\chi^2 = 10.10$, $p < .001$).

[22] A related point of interest further supporting our choice is that we see a relative slow-down on the second word in both of the felicitous conditions: In both cases, the second word in the sequence is read more slowly than when it appears in the first position in the other condition: *wieder* is significantly faster in the *wieder nicht* condition: 449 ms vs. 610 ms (RES-1: $\beta = 161.7$, $SE = 43.29$, $t = 3.74$; $\chi^2 = 12.82$, $p < .001$); and *nicht* is significantly faster in the *nicht wieder* condition: 514 ms vs. 587 ms (RES-1: $\beta = 77.87$, $SE = 37.54$, $t = 2.07$; $\chi^2 = 4.19$, $p < .05$). This

As a final measure, we also looked at the self-paced reading times in the felicitous conditions. Reading times in the *nicht wieder* condition (1124 ms) were marginally higher than in the *wieder nicht* condition (1036 ms) when looking at both words in the {*wieder nicht*} sequence (RES-1: $\beta = 81.26$, $SE = 45.72$, $t = 1.78$; $\chi^2 = 3.15$, $p < .1$) as well as pooling these together with the verb (1625 ms vs. 1788 ms) (RES-1: $\beta = 133.97$, $SE = 78.42$, $t = 1.708$; $\chi^2 = 2.90$, $p < .1$).[23] This suggests that even in the felicitous conditions, greater effort was involved in reading when *wieder* was embedded under *nicht*.

3.3.4 Discussion

The results from the stops-making-sense experiment provide a perspective on the time-course with which the judgments caused by the infelicity of the presupposition of *wieder* arise. In line with our proposed interpretation of Experiment 1a, we found that it takes longer to reject the materials based on a contextual clash with the presupposition of *wieder* when it is embedded under negation. This provides direct support for the PTT Hypothesis. Note that this is quite independent from the issue of whether local interpretations are in principle available or not, as the main point we presently care about is when the global interpretation becomes available. And since rejection takes longer when projection is necessary to arrive at the (global) interpretation that forms the basis of the judgment, we have good evidence for a delay in the availability of this interpretation.[24]

The reading times from the felicitous conditions provide additional suggestive evidence that the embedded *wieder* condition is more effortful, with somewhat elevated reading times on the *nicht wieder (+Verb)* sequence (compared to *wieder nicht (+Verb)*). Finally, the rejection rates for the *wieder nicht* and *nicht wieder* conditions did not differ from one another significantly, which further corroborates the results from our rating study (Experiment 1b). If local accommodation was widely available in the *nicht wieder* condition, we would expect a lower rejection rate there.

3.4 General Discussion of Experiments 1a-c

The results from the series of experiments reported above suggest that presupposition projection requires extra time in processing, thus supporting the PTT Hypothesis. In particular, we find evidence for rapid availability of unembedded presuppositions, while global interpretations of presupposition triggers introduced

suggests that the main burden of processing of the {*wieder nicht*} sequence arises on the second word, regardless of their order.

[23] Looking at just the verb, there was a numerical difference in the same direction (664 ms vs. 589 ms) which did not reach significance (RES-1: $\beta = 77.01$, $SE = 51.32$, $t = 1.50$; $\chi^2 = 2.24$, $p < .15$).

[24] For possible alternative perspectives on the delay, see section 5 below.

in embedded environments are only available with a delay. The three experiments reported each contribute a unique piece to the overall story:

In Experiment 1a, we found immediate increases in reading times when the presupposition of unembedded *wieder* was inconsistent with the context. However, no such effects were found when *wieder* was embedded under negation. The immediate effects in the unembedded condition replicate and enhance the findings in Schwarz (2007) and Tiemann et al. (2011), where unembedded presuppositions that were inconsistent with the context also resulted in processing difficulties, as reflected in self-paced reading times. They provide the most detailed evidence to date on the online time course of presupposition interpretation based on the fine-grained temporal resolution of eye tracking (along with recent results from the visual world paradigm; see Romoli et al., 2015; Schwarz, 2015b).

Responses by participants in the rating task in Experiment 1b showed that the ultimate interpretation of *wieder* does indeed involve a global interpretation of the presupposition, which requires projection: contexts that are inconsistent with such an interpretation, but not with a local accommodation interpretation, were judged to be less acceptable than contexts that are consistent with it. This result is in line with the assumption standardly made in the theoretical literature that global interpretations of presupposition triggers in such embedded environments are the default. However, ratings for the *nicht wieder* condition were somewhat lower than for *wieder nicht*, which could be seen as an indication that the former are more difficult to process.

Experiment 1c involved categorical responses by participants, and materials with contexts inconsistent with the presupposition were rejected as frequently for the embedded condition as for the unembedded one, providing further support for global interpretations being the dominant ones for embedded *wieder*. More crucially, while the eye tracking results on their own only support the notion of a delay for presupposition projection indirectly (since there was no delayed effect of inconsistent contexts, but rather no effect at all), the rejection time results from the stops-making-sense task in Experiment 1c provide more direct evidence along these lines, as it took longer to reject a text in inconsistent contexts in the *nicht wieder* condition than in the *wieder nicht* condition. Finally, the reading times from the felicitous context conditions in this experiment provided further suggestive evidence that the *nicht wieder* condition involved additional effort, based on marginally increased reading times on the combined *nicht* and *wieder* regions (both including and excluding the verb).

Taken together, the results from these experiments support the conclusion that presupposition projection comes with a processing cost, and thus are consistent with the PTT Hypothesis but not the IGI Hypothesis. This result fits quite naturally with the theoretical properties of DRT, whose additional layer of structured discourse representation involved in projection introduces additional complexity that could be mapped quite directly onto corresponding processing cost. A theory like Dynamic

Semantics, which aligns more naturally with the IGI Hypothesis, does not provide as straightforward an explanation for our findings, and would need to be supplemented by additional assumptions to be reconciled with the data. We'll turn to a more detailed discussion of theoretical implications of the results in section 5.

3.5 Interim Assessment

The evidence from the studies reported so far provides intriguing support for the PTT Hypothesis, with potentially important implications for presupposition theory. However, there still are various limitations on the generalizability of the results. The extent to which projection was required was limited to one level of embedding, and only involved one type of embedding expression, namely negation. Furthermore, we only were able to detect effects based on the representation of the presupposed meaning indirectly, by presenting presuppositions in infelicitous contexts. The second experiment, discussed in the next section, aims to address these issues and thereby broaden the generality of the overall result.

4 Experiment 2: Presupposition Projection in Conditionals

To further test the hypothesis that presupposition projection is an effortful process, the materials from Experiment 1a were modified so as to include a conditional, which provides an additional layer of embedding. Another change from the previous experiment was that all of the discourses had a felicitous interpretation, though with variation in the location where support for the presupposition was introduced. This provides a broader perspective on how projection is realized in processing, and allows us to compare different processing accounts of projection-related effects in greater detail. The effects in Experiments 1a-c were based on a single step of projection, and detected in light of clashes with the context. But if projection in general, and perhaps the length of the projection path in particular, are what is behind the results of those experiments, then we should also find projection effects in felicitous contexts when we vary the distance (in terms of the hierarchical discourse structure) between the presupposition trigger and the clause supporting it. Experiment 2 thus provides a more general test of our proposed explanation of the results from Experiments 1a-c.

4.1 Methods

Design & Materials Our items were created according to the following pattern: the presupposition of the target sentence was either supported by the global context sentence or by the local *if*-clause. This corresponded to a manipulation of whether negation was present in the context sentence or in the antecedent of the

Condition	Context	Location	Firstword
a	I	local	*wieder*
b	I	global	*nicht*
c	II	global	*wieder*
d	II	local	*nicht*

Table 6: Overview of Conditions and Factors

conditional. As before, we manipulated the order of *wieder* and negation in the target sentence, which in this case was the consequent of the conditional. This yielded a 2 × 2 design, which could be characterized by pairs of factors in various ways. Table 1 provides an overview of the different possible groupings based on what pairs of factors we consider. The details of the variation in the materials will be discussed in more detail in connection with the theoretical analyses in section 4.2.

(16) *Tina war letzte Woche* {***(I)*** *∅ / **(II)** nicht*} *Schlittschuhlaufen. Wenn sie*
Tina was last week {∅ / not} ice-skating. If she
gestern {***(I)*** *nicht / **(II)** ∅* } *Schlittschuhlaufen war, dann...*
yesterday {not / ∅} ice-skating was, then...

(17) *...geht sie heute bestimmt* {***(WN)*** *wieder nicht / **(NW)** nicht wieder*}
...goes she today certainly {*again not / not again*}
Schlittschuhlaufen.
ice-skating.

Participants and Procedure 24 items with variations of the illustrated pattern were created, each with 4 versions for the four conditions. 32 native speakers of German from the University of Tübingen community participated in the experiment. Participants were split into 4 groups, where each participant saw 6 of the sentences per condition, providing us with a balanced number of data points from all conditions for each item and participant. There were 50 filler sentences from other, unrelated experiments. Participants read all sentences on a computer screen while we recorded their eye movements with an EyeLink 1000 eye tracker. Half of the items were followed by a simple yes/no question to ensure that participants were reading the materials for comprehension.

4.2 Predictions

Before turning to the results, let us spell out the how the relevant theoretical accounts analyze our materials and how this relates to the two processing hypotheses under consideration. To do so, we have to introduce the way they handle conditionals in general, and presupposition projection in conditionals in particular.

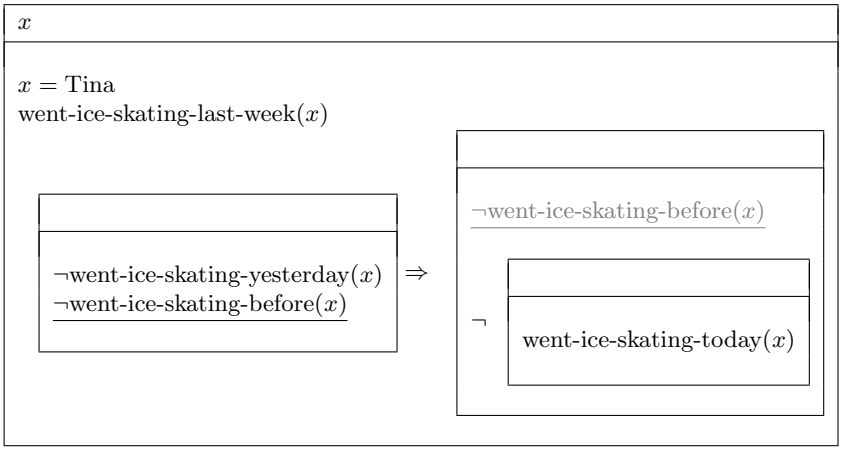
4.2.1 Costly Projection Step-by-Step

We formulated a processing hypothesis that presupposition projection takes time, which was based on the representational complexity involved in a DRT account. In this section we will develop this hypothesis further in terms of the DRSs for the respective sentences.

Operators such as conditionals introduce new embedded DRSs into the structure in DRT. Depending on where the presupposition is introduced and where a suitable antecedent can be found, different projection path lengths come about. Conditionals introduce two sub-DRSs, connected by an arrow. For presuppositions introduced in the consequent, the first place to look for support for the presupposition is the antecedent of the conditional, followed by higher levels. Let us take a look at the DRSs for our four conditions. In condition (a), illustrated in Figure 2, the presupposition that there is an earlier time where Tina did not go ice-skating is introduced in the top-level box of the consequent and satisfied locally in the antecedent of the conditional.[25] Resolution of the presupposition then only involves one step, yielding a projection path length of 1. In condition (b), illustrated in Figure 3, on the other hand, the presupposition that Tina did go ice-skating before is introduced inside of the negated box in the consequent and resolved at the top-most level, based on the information introduced through the context sentence. Given standard DRT-assumptions about the path followed in searching for an antecedent, this involves three projection steps. Turning to condition (c) (see Figure 4), we again have the presupposition of the *wieder nicht* order, as in (a), but this time, given Context II, it is resolved by the context sentence at the top-most level, thus rendering a projection path length of 2. Finally, condition (d) (see Figure 5) has the same presupposition as (b), introduced in the scope of negation inside of the consequent, but this time it is resolved locally in the antecedent of the conditional. This also yields a projection path length of 2.

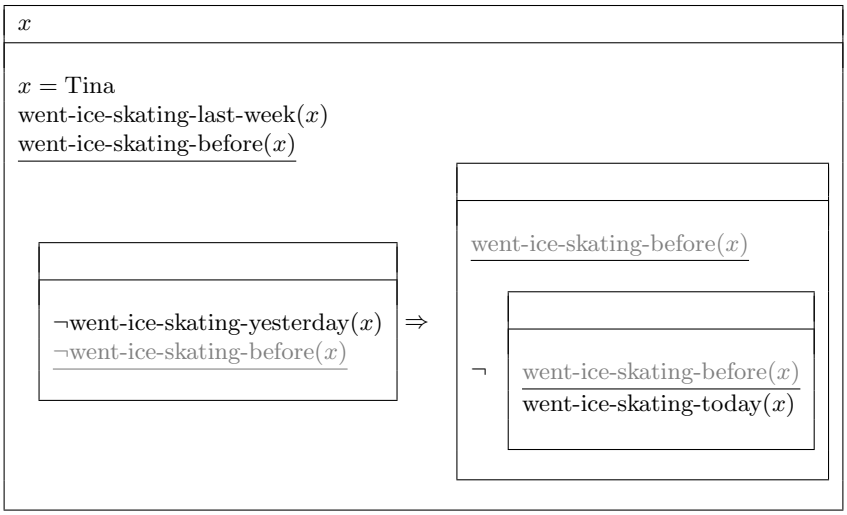
Looking at the distribution of the projection path lengths, we see that an interesting pattern emerges. For Context I, we get a difference based on the location where support for the presupposition is introduced (corresponding to a difference between the *nicht wieder* and *wieder nicht* conditions), with a longer projection path for the global condition (3) than for the local condition (1). Context II, in contrast,

[25] Note that there is a crucial difference between this type of local resolution on the hierarchically closest level and local accommodation in the position where the presupposition was originally introduced; see the general discussion for further details.



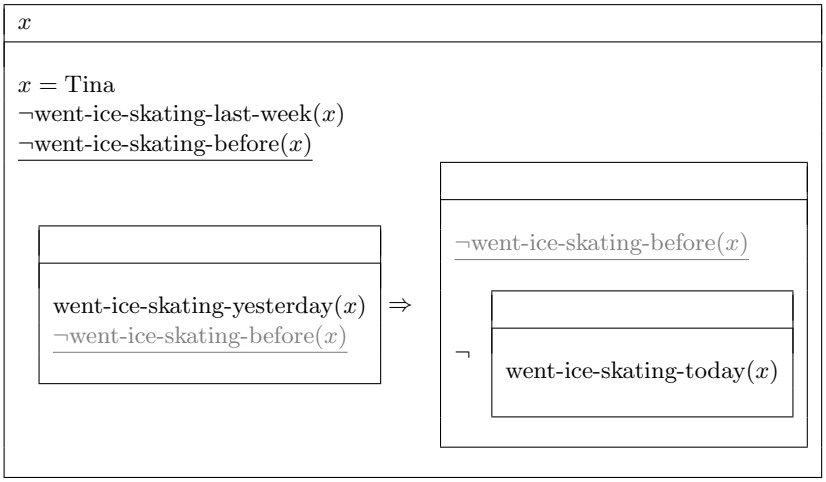
*Tina went ice-skating last week. If she did **not** go ice-skating yesterday, then she **again won't** go ice-skating today.* **Projection Path Length = 1**

Figure 2: **Condition a:** Context: I, Location: local, Firstword: *wieder*.



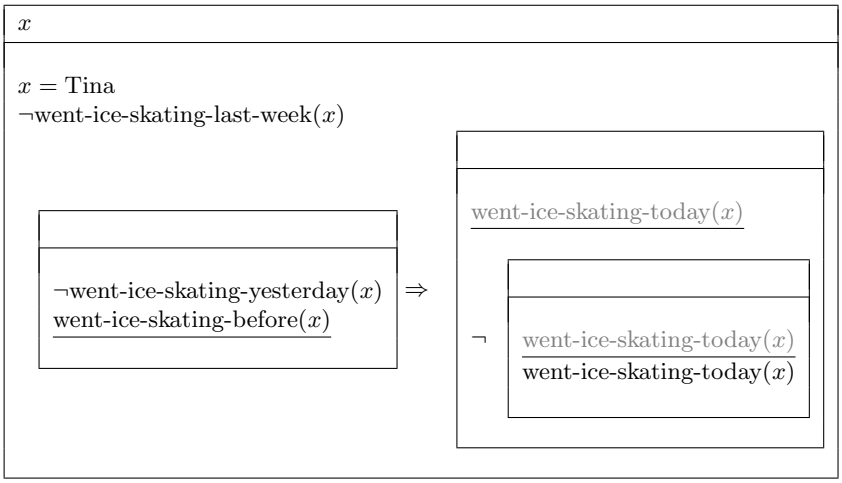
*Tina went ice-skating last week. If she did **not** go ice-skating yesterday, then she **won't** [go ice-skating **again**] today.* **Projection Path Length = 3**

Figure 3: **Condition b:** Context: I, Location: global, Firstword *nicht*



*Tina **didn't** go ice-skating last week. If she went ice-skating yesterday, then she **again won't** go ice-skating today.* **Projection Path Length = 2**

Figure 4: **Condition c:** Context: II, Location: global, Firstword: *wieder*



*Tina **didn't** go ice-skating last week. If she went ice-skating yesterday, then she **won't** [go ice-skating **again**] today.* **Projection Path Length = 2**

Figure 5: **Condition d**: Context: II, Location: local, Firstword: *nicht*

does not give rise to such a difference, as both the global and local conditions (corresponding to *wieder nicht* and *nicht wieder* respectively) have a projection path length of 2. This comes about because of the additional embedding introduced by negation. For the *nicht wieder* condition, this results in two steps, first out of the negation box into the main consequent box, then to the box for the antecedent of the conditional. For the *wieder nicht* condition, on the other hand, there also are two steps, but different ones: from starting out in the top-most box of the consequent, there's one step to the antecedent, and another to the global level.

Fleshing out the PTT Hypothesis by attributing processing costs to each individual projection step (as construed in DRT), we then expect longer projection paths to correspond to greater cognitive efforts, as reflected in increased reading times. For our materials, the DRT projection path distribution then leads us to expect an interaction between **Context** and **Location**, with a difference between the (a) and (b) conditions, but no difference between the (c) and (d) conditions, as illustrated in the left panel of Figure 6 (a main effect of **Location** may also arise based on these predictions, though it would be dominated by the interaction).

4.2.2 Parallel Access to Global and Local Levels

In 2.4 we formulated the IGI Hypothesis based on Dynamic Semantics. Even though this is not a processing theory in its own right, we argued that its theoretical features (without considering additional, more processing related complications) suggest that the global interpretation of a presupposition under negation should be available immediately. In this section we will spell out how the IGI Hypothesis would deal with the more complex embedding structure used in this experiment. As before, we will illustrate this hypothesis by the means of Dynamic Semantics.

A simple view of conditionals in Dynamic Semantics assumes the equivalent of material implication as their meaning contribution, i.e., *if p, then q* is assumed to be true unless p is true and q is false. To do this dynamically, and to account for presupposition projection, this is done in a slightly more round-about way, namely by removing those worlds from c where p is true and q is false. The update procedure does this by first removing the combination of c, p, and q from the combination of c and p, and then removing the resulting set from the initial context c, as can be seen in the formal formulation of the context change potential of a conditional in (18-a):

(18) a. $c+ \text{ If } p, q = c - ((c + p) - ((c + p) + q))$
b. defined iff $(c + p) + PSP_q = (c + p)$

As far as presuppositions in the consequent are concerned, what is important is that the presuppositions of q have to be met in the context that it is added to, but in this case, this context turns out to be the combination of $c+p$. That is, the context update for a conditional can only be performed if the global context c updated with the antecedent of the conditional p entails the presuppositions introduced in the

consequent of the conditional q. This means, however, that it does not make a difference whether the presupposition is supported by the global context (here: the preceding sentence) or the local context (here: antecedent of the conditional) - at least not without further assumptions, which we consider below. The fact that this holds indeed is a crucial piece of the Dynamic Semantic account of presupposition projection, as this allows presuppositional support to be introduced a) in the global context, b) in the antecedent of the conditional, or c) by both combined.

Extending the IGI processing hypothesis from above, this theoretical picture suggests that we would not expect any differences based on the depth of embedding and the distance of the presupposition trigger from the supporting information, as long as the global context is part of the relevant local context. This is not to say that a dynamic semantic account is strictly incompatible with any potential differences, but the underlying semantic procedure alone does not lend itself to explaining such differences. Possible additional factors compatible with Dynamic Semantics, as well as alternative processing implementations for it, are considered in the following section, as well as in section 5.2.2.

Considering other potential processing effects from the perspective of a dynamic account, the most plausible factor at play might be negation, in particular the hypothesis that a negated presupposition is harder to process than an unnegated presupposition, both on intuitive grounds, and based on the more complicated update steps involved. If we take r in (19) to be the unnegated version of the presupposition in our materials, the underlying mechanisms to check the presuppositions of the *again not* and *not again* variants would involve (20-a) and (20-b) respectively.

(19) r = Tina had been ice-skating before

(20)
a. $PSP_q = \neg r:$ $\quad c + \neg r = c? \quad \approx \quad c - (c + r) = c?$
b. $PSP_q = r:$ $\quad c + r = c?$

Hypothesizing that negation adds extra complexity would then suggest that the *again not* sentences should be harder to process than the *not again* sentences, based on the presupposition evaluation process. However, the point about the evaluation of negation requiring extra resources presumably should be applied to the processing of non-presupposed content as well. In our case, this would suggest that negation in the antecedent is relevant as well, which would increase the the relative difficulty of the sentences in Context I, but not Context II, as illustrated in the update process involved in assessing the definedness of the overall context update for our four conditions below.

(21)

	Context I		Context II
a)	$(c' + \neg q) + \neg \mathbf{r} = c' + \neg \mathbf{q}$	c)	$(c' + q) + \neg \mathbf{r} = c' + q$
b)	$(c' + \neg q) \ + \ r = c' + \neg \mathbf{q}$	d)	$(c' + q) \ + \ r = c' + q$

Factoring negation into the equation along these lines, a dynamic account thus also gives rise to the prediction of a **Location**×**Context** interaction. However, this

interaction is crucially different from the one predicted by a PTT account in that the expected simple effect of **Location** in Context I is in the opposite direction and a simple effect of **Location** is predicted in Context II (see middle panel of Figure 6).

4.2.3 Global vs. Local Hypothesis

However, there is at least one further perspective on the experimental materials that one could consider, namely that what matters is whether the presupposition is supported in the global discourse context or within the sentence. In particular, it could plausibly be assumed that the greater distance involved in the global context case corresponds to an increase in processing cost. This could be motivated in various ways, e.g., by alluding to distance effects for anaphora resolution, since *again* commonly is seen as an 'anaphoric' presupposition trigger. Along the same lines, a processing account based on Dynamic Semantics could assume that context change potentials are initially computed in isolation for the sentence at hand (presumably assuming a trivial place-holder for c for purposes of computation), and then integrated with the actual discourse context c (we flesh this idea out further in section 5.2.2).[26] Similarly, from the perspective of psychological models of discourse processing based on situation models (e.g., Zwaan & Radvansky, 1998), one could allude to the difference between the 'current model' and the 'integrated model', which are stored in different levels of working memory (see section 5.3.2 for a brief discussion of such models).[27]

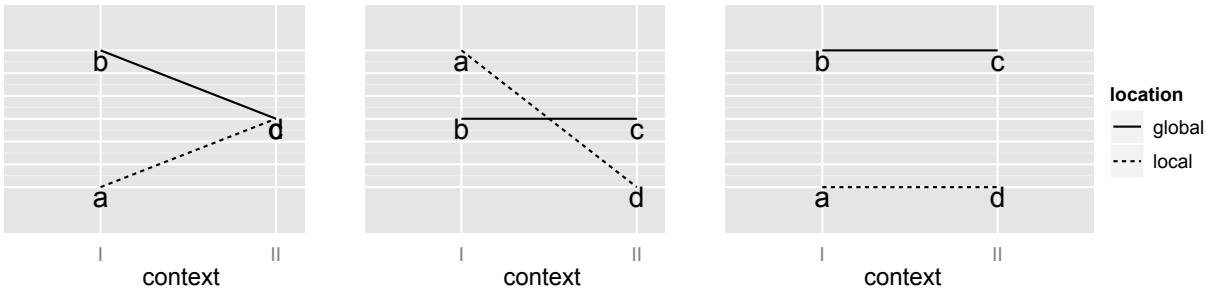
Regardless of which theoretical motivation for this hypothesis we adopt, the key issue for this hypothesis is simply whether the antecedent for the presupposition is introduced in the global context sentence or in the local antecedent of the conditional. This predicts a main effect of **Location**, with no differences based on **Context** (or **Firstword**). This prediction is illustrated on the right side of Figure 6.

4.3 Results

As in Experiment 1a, the primary focus in our analysis were the reading times on the verb following the {*wieder nicht*} sequence in the consequent of the conditional, since this is the point at which the presupposition becomes fully explicit, but we also analyzed reading times on the {*wieder nicht*} sequence itself. The same standard reading measures as in Experiment 1a were calculated, and the analyses proceeded as before, using linear mixed effect models with the maximal random effect structure that would converge, and model comparisons to assess the impact of a given factor on the overall fit of the model. Significant effects were found for regression path duration and total reading times. Re-reading times also gave rise to some significant

[26] Thanks to an anonymous reviewer for raising this possibility.

[27] Some elaborations of this type of hypothesis are considered in section 5.2.2.



Projection-Takes-Time Immediate-Global Access Global vs. Local

Figure 6: Processing Predictions

effects, but the number of data points was very small for these, so we focus on the former two in the presentation of our results. Trials where any blinks occurred while looking at the region of interest were removed prior to analysis. Data points that deviated by more than three standard deviations from the mean of their condition were excluded from the analysis as well.

The mean reading times for the two measures to be discussed here are summarized in Table 7, and the general pattern of results is illustrated in Figure 7. The main set of statistical analyses were carried out using **Location** and **Context** as

		Context I		Context II	
	firstword	*wieder*	*nicht*	*wieder*	*nicht*
	Location	local	global	global	local
	Condition	a	b	c	d
verb	Reg. Duration	340	411	375	385
	Total	396	493	432	427
{*wieder nicht*}	Reg. Duration	341	413	381	389
	Total	400	489	435	429

Table 7: Reading times on Verb and *wieder nicht* sequence (in ms)

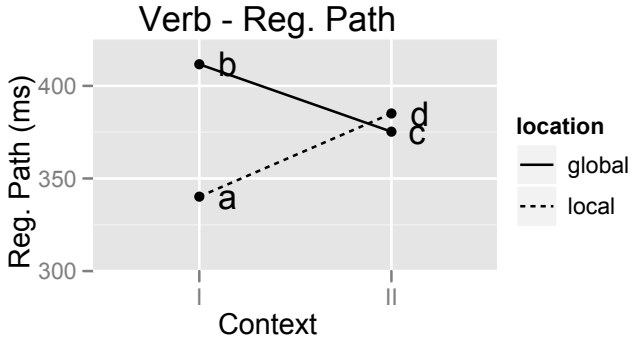


Figure 7: Summary of Regression Path Duration for the verb (in ms)

factors to assess the predictions of the different accounts laid out above. The central result is that there is a significant difference between conditions a and b, but no reliable difference between c and d, i.e., the local vs. global distinction mattered in Context I (where *wieder nicht* was paired with local and *nicht wieder* with global), but not in Context II (with the reverse pairing). Statistically, this was supported by a significant interaction between context and location.

For Regression Path Duration, the interaction between Context and Location significantly improved overall model fit (RES-2: $\beta = 78.39$, $SE = 35.92$, $t = 2.18$; $\chi^2 = 4.6112$, $p < .05$). The same held for total reading time (RES-4: $\beta = 85.11$, $SE = 35.7$, $t = 2.38$; $\chi^2 = 5.65$, $p < .05$). For the latter, there also was a significant main effect of Location (RES-2: $\beta = 53.86$, $SE = 19.18$, $t = 2.81$; $\chi^2 = 6.67$, $p < .01$), though note that this was dominated by the cross-over interaction. For regression path duration, this effect approached, but did not reach, marginal significance (RES-2: $\beta = 31.96$, $SE = 19.35$, $t = 1.65$; $\chi^2 = 2.69$, $p < .11$). The interactions were driven by the difference between conditions (a) and (b), both for regression path duration (RES-1: $\beta = 70.94$, $SE = 28.97$, $t = 2.45$; $\chi^2 = 5.69$, $p < .05$) and total duration (RES-2: $\beta = 86.01$, $SE = 37.78$, $t = 2.28$; $\chi^2 = 11.29$, $p < .001$). No significant differences emerged between conditions (c) and (d).

Comparable results emerged for these two reading time measures when looking at the {*wieder nicht*} region: there was an interaction for both regression path duration (RES-3: $\beta = 71.06$, $SE = 31.19$, $t = 2.28$; $\chi^2 = 5.18$, $p < .05$) and total reading time (RES-3: $\beta = 77.46$, $SE = 33.80$, $t = 2.29$; $\chi^2 = 5.23$, $p < .05$), as well as a main effect of Location for the latter (RES-2: $\beta = 50.27$, $SE = 19.39$, $t = 2.59$; $\chi^2 = 5.91$, $p < .05$) and a marginal one for the former (RES-3: $\beta = 32.00$, $SE = 16.24$, $t = 1.97$; $\chi^2 = 3.79$, $p < .1$). As with the verb-region, the interaction

was driven by simple effects between conditions (a) and (b), both for regression path duration (RES-1: $\beta = 68.09$, $SE = 23.87$, $t = 2.85$; $\chi^2 = 7.79$, $p < .01$) and total duration (RES-1: $\beta = 88.94$, $SE = 24.57$, $t = 3.62$; $\chi^2 = 12.52$, $p < .001$).

Given the nature of our factors, we conducted statistical analyses for several other pairwise comparisons. Within the global conditions, we found a significant difference for total times ((b) > (c)), both on the {*nicht wieder*} (RES-3: $\beta = 52.68$, $SE = 24.22$, $t = 2.18$; $\chi^2 = 4.73$, $p < .05$) and verb regions (RES-2: $\beta = 61.62$, $SE = 26.81$, $t = 2.30$; $\chi^2 = 5.189$, $p < .05$). A marginally significant difference in regression path duration was found between the local conditions ((d) > (a)) on the {*nicht wieder*} region (RES-1: $\beta = 40.43$, $SE = 21.90$, $t = 1.85$; $\chi^2 = 3.41$, $p < .1$). Furthermore, there were differences between the two *nicht wieder* conditions for total duration ((b) > (d)), both for the *nicht wieder* region (RES-2: $\beta = 65$, $SE = 25.27$, $t = 2.57$; $\chi^2 = 61.6$, $p < .05$) and for the verb region (RES-4: $\beta = 72.45$, $SE = 25.80$, $t = 2.81$; $\chi^2 = 7.66$, $p < .01$). Finally, the *wieder nicht* conditions differed marginally from one another ((c) > (a)) on the *wieder nicht* region for regression path duration (RES-1: $\beta = 36.99$, $SE = 22.13$, $t = 1.67$; $\chi^2 = 2.79$, $p < .1$).

Altogether, the pattern of the results very closely resembles the predictions of the PTT Hypothesis. The greatest and clearest difference in reading times is found for the two conditions with the greatest difference in projection path length, namely (a) and (b), whereas no difference is found between (c) and (d), counter to what we would expect based on the linear distance hypothesis. Furthermore, for total reading times, we also find differences based on the level of the **Location** factor for both *wieder nicht* and *nicht wieder*, with longer reading times in the global conditions ((b) > (c)). There also are some significant differences between the local and global contexts for the *nicht wieder* condition ((b) > (d), as well as a marginal one for the *wieder nicht* condition ((c) > (a)). These results correspond surprisingly well to the pattern expected based on DRT projection path length. Based on this observation, we conducted an additional follow-up analysis, in which we tried to model the data using only projection path length as a (numerical) predictor. Both for regression path duration and total time, this yielded a significant effect of projection path length ($t's > 2.8$). Based on model comparisons with the interaction analysis above, the resulting fit was not significantly different for the two analyses. In other words, projection path length alone was as good a predictor of reading times as the **Location**×**Context** interaction.

4.4 Discussion

The results for total reading time and regression path duration mirror the predictions of the PTT hypothesis based on DRT projection path length remarkably well, and indeed the data can be modeled adequately by just considering projection path

length as a predictor.[28] This is in line with our proposed interpretation of the results from the first set of experiments, based on the idea that presupposition projection involves representational complexity of some kind, and that representational complexity has real cognitive correlates. As in the first set of experiments, an IGI Hypothesis based on Dynamic Semantics either doesn't predict any differences at all, or it predicts an interaction of a very different nature, if we assume that negation introduces processing complexity. Similarly, the predictions of the Global vs. Local Hypothesis are not borne out in the data either, as they only involve a main effect based on whether the presupposition is supported by the context sentence or the antecedent of the conditional. However, this did not make a difference if one kept the overall projection path length constant by switching from the *wieder nicht* (c) condition to the *nicht wieder* condition (d). What rather seems to matter in terms of processing effort is the length of the projection path overall, not the absolute location at which support for a presupposition is introduced.

5 General Discussion

5.1 Summary

Our starting point were two processing hypotheses about presupposition projection in online processing, the PTT Hypothesis and the IGI Hypothesis, which in turn took DRT and Dynamic Semantics as their respective theoretical starting points. We presented four experiments which aimed to shed light on the impact of embedding on the online interpretation of the presupposition of German *wieder*. In Experiment 1a, we found immediate reflexes in reading time measures when unembedded *wieder* did not match the context, but no corresponding delays were found for embedded *wieder*. Our proposed explanation for this was that in the latter case, the global presupposition is not immediately available, in line with the PTT Hypothesis. The results from the rating task in Experiment 1b confirmed that participants generally do perceive the infelicity for both the *wieder nicht* and *nicht wieder* versions in the respective contexts, and the rate of stops-making-sense judgments in Experiment 1c furthermore was comparable for both types of conditions. It thus does not seem possible to explain the reading time results in terms of a general availability of local accommodation. Tying together these judgments with the time-course of interpretation, we also looked at the rejection times in the stops-making-sense task. Consistent with the hypothesis that global presuppositions of embedded presupposition triggers only become available with a delay compared to unembedded ones, we found longer rejection times for the *nicht wieder* condition. Looking at word-by-word reading times in the felicitous conditions, we also found tentative evidence for slow-downs

[28] It's worth noting that there's a difference in terms of what reading time measures exhibit the effects in the two eye tracking experiments. Our tentative explanation for this is that the types of sentences in experiment 2 are somewhat more complex, and that we are only dealing with felicitous texts.

in the *nicht wieder* condition, even when the presupposition was supported by the context. Experiment 2 attempted to broaden the evidence for our interpretation of the first set of studies by looking at additional embeddings, and further variation in the location of support of the presupposition. The results neatly match the predictions of the PTT Hypothesis, which fits naturally with a correspondence between processing effort and the level of representational steps that have to be taken to arrive at the final interpretation.

5.2 Theoretical Implications: Potential Additional Factors and Related Results from the Literature

Taken together, the results from the reported experiments are in line with the PTT Hypothesis, but are not accounted for by the IGI Hypothesis. If the theoretical motivation behind the former hypothesis is indeed on the right track, then the results can be seen as lending experimental support to the representational complexity involved in presupposition projection as posited by DRT. However, it could also be that there are other factors behind the experimental results, which in turn could be entirely compatible with the alternative semantic perspective of Dynamic Semantics (or other theoretical proposals; see below). Furthermore, alternative processing implementations of dynamic semantics may lead to different predictions. In the following, we review a variety of possible candidates for such alternative explanations of our data, as well as related results from the literature that might seem to conflict with our findings. Note that one shared property of all of the alternatives considered below is that they primarily posit an additional factor that would predict a main effect. The observed interactions then have to be seen as resulting either from the interplay of two factors introducing main effects, or from a floor effect coming into play with a main effect. This is in contrast with the interpretation of the data we offered above, which accounts for the observed interactions in a unified manner. While neither approach necessarily is conceptually superior to the other, the former type of account naturally has more assumptions going into it, as each factor should ultimately be independently motivated.

5.2.1 Potential Advantages of *Nicht Wieder* over *Wieder Nicht*

We begin with an alternative interpretation of the eye tracking results from Expt 1a, suggested to us by Judith Degen (p.c.): it seems to be the case that the form *nicht wieder* is more frequent than *wieder nicht*.[29] If higher frequency forms give rise to decreased reading times, which furthermore might extend to a following word (in our

[29] According to a quick search in the Guttenberg corpus performed by Degen, the sequence 'NOUN nicht wieder' yields 1395 occurrences, as opposed to 95 for the sequence ''NOUN wieder nicht'. Google searches for the bigrams 'nicht wieder' and 'wieder nicht' yield about twice as many hits for the former. While none of these measures may be perfect, the general picture of a frequency asymmetry seems relatively clear.

case the verb), then a possible explanation of the lack of felicity effects in reading times in the *nicht wieder* condition could be based on a floor effect: Fixations can only be so short, and a strong frequency effect might bring the infelicitous condition down to a minimum, erasing any further advantage based on felicity. While such an account is entirely consistent with the basic results from Expt 1a, the experimental data on the whole is not.

The floor-effect account crucially rests on the assumption that *nicht wieder (+Verb)* is easier and faster to process than *wieder nicht (+Verb)*, which makes a straightforward prediction for what should happen in the absence of infelicity: reading times for the latter should be greater than for the former. This prediction is not borne out in our data. In Experiment 1a, total reading times for the felicitous *nicht wieder* condition are greater than for the *wieder nicht* one. The same holds for combined self-paced reading times on {*nicht wieder*} and {*nicht wieder*} *+Verb* in the felicitous conditions of Experiment 1c. Similarly, there were simple effects based on the order of *nicht* and *wieder* for total times and regression path duration on {*wieder nicht*} and the verb in Experiment 2, both in the local and global contexts. Relatedly, also note that the ratings for the *nicht wieder* conditions in Experiment 1b were lower across the board, which could be seen as a further indication of increased complexity and corresponding processing difficulty. Finally, a floor-effect account of the results from Experiment 1a would have to explain why there is no comparable floor-effect in Experiment 1c, where stops-making-sense judgments in the *nicht wieder* condition were slower than in the *wieder nicht* condition. Apart from these aspects of the overall data that pose empirical challenges to a floor-effect account, it is not clear whether the conceptual motivation of such an account survives closer scrutiny.

In particular, the effect that we do find in the unembedded condition of Experiment 1a is due to the inconsistency of the presupposition with the preceding context. We have interpreted this as an indication that the presupposition is immediately available and evaluated relative to the context. This allows for immediate detection of the infelicity, and it is the resulting clash of the context that gives rise to the observed delay. If the *nicht wieder* condition were easier to process due to its higher frequency, as posited by the floor-effect account, then the relevant presupposition should be available even more easily and quickly. But then the inconsistency should be immediately detected, as in the unembedded case, and give rise to the same perception of infelicity, with a corresponding slow-down in this condition. In other words, the proposed speed-up in processing *nicht wieder* is entirely orthogonal to the slow down due to infelicity: the former should be present in both the felicitous and infelicitous versions, the latter only in the infelicitous one. We find it hard to see how a floor-effect on the former should wipe out any independent effect of the latter. Be this as it may, given the relative slowness of the *nicht wieder* felicitous conditions, we do not find this alternative account convincing when considering the full set of data from our experiments.

Another related possible explanation for the eye tracking results from Experiment 1a based on order effects was pointed out by Jesse Snedeker (p.c. at CUNY 2012) is the following. It might be the case (perhaps also due to the frequency asymmetry) that participants get used to dealing with the infelicity more quickly in the *nicht wieder* condition. The reading times in the infelicitous conditions then would be expected to start out at comparable levels on the first couple of trials, but decrease more quickly for the *nicht wieder* condition. Including order as a factor in a linear regression analysis does not support this possibility. While the only significant effect is the Firstword*Felicity interaction, reading times in the infelicitous *nicht wieder* condition remain at a stable low level throughout, whereas those for the infelicitous *wieder nicht* condition decrease numerically over the course of the experiment. From the perspective of our interpretation of the data, this suggests that participants initially stumble badly over the infelicity in the unembedded condition, but get somewhat used to it over the course of the experiment.

5.2.2 Advantages for *Wieder nicht* and Local Contexts

Potential Advantages of *Wieder Nicht* over *Nicht Wieder* Yet another potential alternative approach to the results from Experiment 2, originally suggested to us by Colin Phillips and Jeff Lidz (p.c.), is based on a possible main effect due to an independent advantage of *wieder nicht* over *nicht wieder* (i.e., the exact opposite from the previous possibility). There are various possible motivations for such an advantage. For example, having negation in the scope of *wieder* and adjacent to the verb might involve greater predictability of the verb (e.g., based on a more limited search space for a 'negative' presupposition), or involve decreased memory costs because the meaning of negation can be integrated with the verb immediately and on its own, rather than in parallel with the presupposition of *wieder*. Alternatively, the fact that *wieder* in the scope of negation in principle is associated with two possible interpretations (the global and local readings discussed above), might give the alternative *wieder nicht* variant an advantage due to the simpler set of interpretive options.[30]

Such an advantage of *wieder nicht* over *nicht wieder* could also be derived from a revised processing implementation of dynamic semantics (suggested to us by Raj Singh, p.c.). Up to now, we followed what seems to us like the most basic approach one could take in this regard, based on the idea that everything that factors into the local context relative to which a presupposition is evaluated contributes equally and in parallel, and that evaluation relative to the same local context corresponds to equivalence in processing. However, a variation of this picture might help to capture at least part of the results we attributed to differences in levels of embedding and presuppositional support. It starts from the natural observation that updates with a negated sentence can be considered as more complex than updates with

[30]Thanks to Uli Sauerland (p.c.) for this suggestion.

affirmative sentences. This alone wouldn't help much, since our materials always contained negation and merely varied the relative positioning of *again*. However, one could add that evaluating a presupposition - which we represented procedurally by checking whether addition of the presupposed proposition does not alter the local context - as part of an overall more complex evaluation procedure induces additional cost. Consider the simple illustrations in (22) and (23). According to the present hypothesis, carrying out the very same step might be more effortful in ((22-a-ii)), as compared to ((23-b)), because it takes place within a more complex overall sequence of steps. This might involve a greater memory load in carrying out step ((22-a-ii)). (A parallel point applies for presuppositions in conditionals, as compared to unembedded ones.)

(22) Computation of negation
- a. $c + \neg S_p = c - (c + S_p)$
  - (i) $c + S_p$
  - (ii) check $c + p = c$
  - (iii) if so: $c + S = q$
- b. $c - q$

(23) Computation of affirmative
- a. $c + S_p$
- b. check $c + p = c$
- c. if so: $c + S$

This would account for the additional complexity introduced by having to assess a presupposition relative to a local context as part of a larger, more complex evaluation procedure, and thus capture the advantage of *wieder nicht* over *nicht wieder*.

While we are not in a position to thoroughly assess the viability of these potential motivations for an advantage of *wieder nicht*, we can consider the general prediction that positing any such factor makes, namely a main effect of our **firstword** factor that should be present throughout. As such, it does not account for the additional effect of varying the location of presuppositional support in Experiment 2, which was independent of of the *wieder nicht* vs. *nicht wieder* variation and led to the various interaction effects we observed. However, it could be combined with another factor, in particular the one encoded in the Global vs. Local Hypothesis from above. We return to this next, and then discuss a possible account of the data in terms of a combination of these two factors.

Global vs. Local Revisited In section 4.2.3 we already introduced the hypothesis that there might be a general advantage of local presupposition resolution within the sentence over global resolution. Such a preference could be motivated based on a number of broadly related considerations. First, *wieder* is commonly regarded as an anaphoric trigger, which thus may exhibit effects common to anaphora resolution. For example, from the perspective of memory load, greater distance plausibly will be associated with greater cost. There's a question here as to what the appropriate distance measure is (e.g., number of words, clauses, nodes at LF), but in any case, it's going to be something on a (more or less) continuous scale. Another perspective

might make more of a binary distinction by drawing a line between sentence-internal information and the sentence-external discourse context.

One reviewer suggests that another variant of a processing implementation of dynamic semantics could yield such a division by initially computing context change potentials of sentences first and only then evaluating them relative to the actual discourse context.[31] This would make it possible to differentiate between introducing presuppositional support in the antecedent and in the context sentence in our materials for Experiment 2. Here's a sketch of a possible implementation of this idea.[32] Assume that context change formulas are evaluated strictly bottom-up (or inside-out), without taking into consideration information from higher nodes, including the actual content of the overall common ground.[33] In other words, occurrences of c (the common ground prior to the utterance of the sentence) below the top-level of the tree are temporarily treated as variables whose value has not been resolved. In that case, it would greatly matter where the support of a given presupposition is introduced, as the point where satisfaction of a presupposition can be determined would vary. Consider the following schematic illustration of the case of a conditional with a presupposition p in its consequent:

(24) Computation of a conditional:

a. $c+$ *if* $S, T_p = c - ((c + S) - ((c + S) + T_p))$
   (i) $c_i + S = q^{c_i}$
   (ii) $q^{c_i} + T_p$
   (iii) check $q^{c_i} + p = q^{c_i}$

b. if so:
   (i) $q^{c_i} + T = r^{c_i}$
   (ii) $q^{c_i} - r^{c_i} = u^{c_i}$
   (iii) $c - u^{c_i} = c - u^{[c_i \rightarrow c]}$

c. if not:
   (i) $q_p^{c_i} + T = r_p^{c_i}$
   (ii) $q_p^{c_i} - r_p^{c_i} = u_p^{c_i}$
   (iii) $c - u_p^{c_i} = c - u_p^{[c_i \rightarrow c]}$
   (iv) check $(c + S) + p = (c + S)$
   (v) if so: $c - u^{[c_i \rightarrow c]}$

Up until reaching the level of the global c, the place holder c_i is used, i.e., no information from the global context is present until the final lines of (b) and (c). If the presupposition p is already entailed by the antecedent S, regardless of the nature of the context c that ultimately comes into play, then it can be determined

[31] Thanks are also due to Philippe Schlenker, who raised a related point.

[32] Thanks to Raj Singh (p.c.) for helpful discussion on this.

[33] It might also be possible to relate this to the notion that given the uncertainty about actual discourse participants' belief states, we might entertain a whole range of possible common grounds at any given point (Beaver, 2001).

early on, in step (a-iii), that the presupposition is satisfied in its local context. Thus, the presupposition can be neglected from this point on, and the evaluation of the entire formula proceeds normally.[34] However, if S alone does not entail p, then p has to be carried along, as it were, (indicated by the subscript p) and re-evaluated in step (c-iv), when the content of c becomes available. Such a strictly inside-out evaluation procedure would thus predict that conditionals with a presupposition in their consequent that requires the support of the global context c (possibly in combination with S) are more complex than parallel cases where the antecedent alone supports the presupposition.[35]

Such a perspective is of course not exclusively available to Dynamic Semantic approaches. It could equally be formulated within DRT, or, for that matter, certain trivalent accounts of presuppositions, as pointed out by another reviewer. The latter can differentiate between the evaluation of presupposition satisfaction at the utterance level, and the sentence-internal compositional impact of presupposition-related information that accounts for projection phenomena (such a view is discussed, for example, in Fox, 2008). Finally, from the perspective of psychological theories of discourse processing based, e.g., on situation models, one can distinguish different level models that are part of the process of integrating information from a new sentence with information from the preceding discourse (more on this in section 5.3.2 below). The general prediction of any such account is going to be that local support incurs less cost than more global support. On its own, this again can't account for the full range of data, but it comes closer to doing so when combined with the previous factor.

Combining these two factors Whichever motivation we go with for either one of the two factors just considered, the two of them together would predict an interaction of **Context** and **Location** comparable to that captured by the interpretation we have put fourth, since the two conditions for which we found large reading time differences involve both a switch from *nicht wieder* to *wieder nicht* and from a local to a global context. So if we looked at the data from the perspective of the factors of **Firstword** and **Location**, we might simply see additive main effects of the two. However, it is not clear that the overall evidence from our experiments provides compelling support for this view. In particular, the results from Experiment 1a do not seem to fully align with it.

If *nicht wieder* in general were at a disadvantage with respect to *wieder nicht*, we would expect corresponding differences to come into play when comparing the infelicitous conditions in Experiment 1a to one another. Depending on whether the

[34] The final step in (b-iii) is intended to suggest that the content of c is supplied for all instances of c_i within u. Note that this is not a fleshed out formal proposal, but merely a suggestive illustration.

[35] As pointed out by Raj Singh (p.c.), it may be fruitful to relate this proposal to dynamic approaches to the proviso problem that appeal to a role for structural factors of one sort or another, e.g., (Heim, 2006; Singh, 2007; Schlenker, 2011; Fox, 2012). Space constraints preclude us from exploring this more fully, but we hope that the idea will be taken up in future work.

alternative account assumed a general effect of infelicity or just for the *wieder nicht* conditions, it would lead us to expect that the infelicitous *nicht wieder* condition would be even worse than the infelicitous *wieder nicht* condition, or at least roughly as bad (assuming the two factors operate on a comparable level in terms of the magnitude of their effects). Furthermore, under the former assumption, we'd expect an effect of context for *nicht wieder* after all. This prediction is not borne out in the results from Experiment 1a above. While a more detailed evaluation of an alternative account along the lines sketched here may be warranted, we thus do not see a strong case in favor of it at present.[36]

An additional point to note is that the respective motivations for the two factors based on processing implementations of dynamic semantics do not seem straightforwardly compatible with one another, since one appeals to an effect of overall complexity at the point of presupposition evaluation, whereas the other puts that very complexity aside initially and adds it back in as the computation proceeds outwards. Therefore, either one of them would need to be combined with one of the other potential motivations for the respective other factor to capture the results in their entirety.

5.2.3 A Floor Effect for Negative Contexts?

An anonymous reviewer suggested another version of an account based on a floor-effect. If we group our test items in Experiment 1 with regard to the context they appear in, we see that target sentences following a 'negative' context sentence (i.e. where no salient event was introduced) were processed faster, both in the felicitous and the infelicitous condition. An alternative explanation for the results in Experiment 1a then could be that participants were less careful in reading the test sentences after a negative context and thus didn't realize the infelicity in the *nicht wieder* case. This raises the question of why subjects rejected the infelicitous *nicht wieder* sentences in Experiment 1c (and rated them lower in 1b), but perhaps the difference in the nature of the task could be held accountable: the explicit judgment task would ensure that subjects paid closer attention in all conditions.

Even though this line of reasoning could account for the results observed in Experiment 1a, it seems unlikely to us that the mere absence of an event in the context sentence would lead people to generally read the subsequent sentences less carefully, and this assumption certainly would need to be further motivated to flesh out this interpretation. However, even more importantly, this proposal also makes predictions for Experiment 2. In particular, if the 'negative' nature of the immediately preceding context leads to less careful reading, one might expect faster reading times on the consequent of the conditional when the antecedent contains negation, i.e., in context I. However, this is precisely where we find reading time effects due

[36] One additional point to note here is that a general advantage of *wieder nicht* over *nicht wieder*, whatever it may consist of, would help to explain the main effect of **firstword** in Experiment 1b.

to embedding of *wieder* under negation, and the overall nature of the results as an interaction is not accounted for on this proposal. As with the previous floor-effect proposal, we thus conclude for the moment that such a factor would require further motivation for one, but even then could not account for the full pattern of results from our experiments on its own.

However, this proposal could be combined with an independent explanation of the effects observed in Experiment 2 to yield better empirical coverage. The same reviewer suggests that perhaps only the *nicht wieder* condition with global support (b) involved a genuine global reading. The other condition that we described as having a global reading, *wieder nicht* in context II (c), could involve local presuppositional support if the *if*-clause is understood exhaustively, thus introducing an interpretation where yesterday (the occasion mentioned in the *if*-clause) was the only time Tina went ice-skating. However, our own native intuitions do not support the presence of an exhaustive interpretation. Even more importantly, it does not seem like such an interpretation can support the presupposition introduced by *wieder* with negation in its scope: without the global context sentence, the relevant version of the sentence does not seem to be fully felicitous, presumably due to the lack of a salient antecedent occasion where Tina did not go ice-skating:

(25) *Wenn Tina gestern Schlittschuhlaufen war, dann geht sie heute*
If Tina yesterday ice-skating was, then goes she today
bestimmt wieder nicht Schlittschuhlaufen.
certainly *again not* ice-skating.

On a related note, Valentine Hacquard has raised the concern whether condition (b) (with *nicht wieder* and global support) differs in felicity from the other conditions, in particular since one equally well could have placed *wieder* in the antecedent clause in that case. It is not clear to us, based on our own intuitions about these sentences, that there is a variation in felicity that goes beyond what can be accounted for in terms of the efforts involved in projection complexity, along the lines argued for above. We therefore leave a more comprehensive assessment of these possible alternative accounts of Experiment 2 for future work.

5.2.4 Relation to Experimental Results on Local Accommodation

As a final point, we'd like to put our results in perspective with another recent study that looks at local accommodation vs. global interpretation of presuppositions:[37] Chemla and Bott (2013) report results from a reaction time study involving truth-value judgments, where 'false'-judgments correspond to global interpretations and 'true'-judgments to local accommodation, which show the former to be faster than the latter. While this may initially seem as if it were the opposite of our findings, we have to take into account the differences between the studies to appreciate that

[37] See also Romoli and Schwarz (2015) for comparable results from a slightly different paradigm.

the results may indeed be entirely compatible. Chemla and Bott used factive verbs such as *realize* to construct sentences such as (26) in the context of a story about alien zoologists:

(26) Zoologists do not realize that elephants are birds.

The verb *realize* introduces a factive presupposition to the effect that its complement clause is true. If this presupposition is interpreted globally relative to the matrix negation, the sentence will naturally be judged to be false. However, if interpreted locally, the sentence essentially conveys that it is false that elephants are birds and zoologists believe so. Such a negated conjunction is, of course, true, since the first conjunct on its own is obviously false. Chemla and Bott compare semantic and pragmatic accounts of presuppositions (namely a dynamic account à la Heim 1983 as compared to a pragmatic account along the lines of Schlenker, 2008a) and argue that the former predict global interpretations to be faster, while the latter predict the reverse. Their finding of slower response times for 'true' answers then seem to support semantic accounts. How does this relate to our results that suggest that longer projection paths lead to increased processing efforts? It fits into the larger picture if we assume that in the Chemla and Bott study, the local responses are based on local accommodation (where the presupposition is interpreted in the very position it was introduced in syntactically). From the perspective of DRT (as well as many other accounts), accommodation is a last resort repair mechanism. More specifically, DRT accounts of presupposition assume that there is a two-phase procedure involved in presupposition evaluation: first, the projection path is followed upwards in the discourse structure in search of an antecedent. If no antecedent is found, global accommodation is attempted, but if this leads to inconsistency, one retreats back down the search path in search for a suitable location for accommodation. On this view, local accommodation is precisely the last stage in this procedure, and it thus is entirely consistent with such an account that local accommodation responses are slower than global ones. In light of this, one might wonder why local accommodation is not more prevalent for our *nicht wieder* sentences in Experiments 1a-c. However, it is quite commonly known that different triggers display varying potential for accommodation in general and local accommodation in particular, so this difference may entirely be due to the fact that *again* is a so-called hard trigger, while factive verbs are soft triggers in the terminology of Abusch (2010). In closing this brief comparison, note that while Chemla and Bott take their results to support Dynamic Semantics, the DRT style analysis we have considered here seems entirely on par with Dynamic Semantics in the relevant respects and thus is equally compatible with their results.

5.3 Broader Connections and Conclusion

In this final section, we briefly point out some connections to broader issues in discourse processing, and present some general conclusions. First, we comment on the impact our results have on the general time-course of presupposition processing. Secondly, we relate our findings to discourse processing models from the psychology literature. The final subsection sums up the central conclusions.

5.3.1 The Time-Course of Presupposition Processing

An important aspect of the reading time results from Experiment 1a was that for unembedded *wieder*, we observed slow-downs due to infelicity as soon as all the content going into the presupposition is made explicit. In particular, we found increases in first fixation duration on the verb. This result is one of the most direct and detailed pieces of evidence for the immediacy of (unembedded) presupposition processing to date (see also Romoli et al., 2015; Schwarz, 2015b, for parallel recent results from the visual world paradigm), and thus contributes an important piece to our overall understanding of how different aspects of meaning unfold in online processing.

Furthermore, note that the delays for embedded *wieder* are highly relevant for the evaluation of pragmatic accounts of presuppositions, as these generally assume all occurrences of presupposition triggers - embedded or unembedded - to involve the same pragmatic reasoning processes. They therefore do not have a general explanation for differences between these contexts at the ready. This is particularly so for accounts such as that proposed by Schlenker (2009), as it is entirely on par with Dynamic Semantics with regard to the inability to differentiate between locations of support for presuppositions, since it evaluates presuppositions relative to all preceding text, without differentiating between, say, the (local) antecedent of a conditional and preceding sentences.

5.3.2 Relation to Discourse Processing Models

As already highlighted in the introduction, presuppositions constitute an ideal phenomenon for investigating questions about discourse processing, as they are intertwined directly both with the immediate, intra-sentential linguistic context as well as the more general discourse context. The difference between the two theoretical proposals we focus on ties in directly with central questions about the processing of linguistic input in the context of a discourse. In particular, there is the general question of the extent to which information introduced linguistically is stored in a structured form. Furthermore, there is the question of whether linguistic input is processed (at least initially) relative to a fairly narrow local window (e.g., consisting of the relevant clause or sentence), or integrated with the overall discourse context right away.

Many processing theories of discourse comprehension crucially involve the notion of a *situation model* (e.g., Zwaan & Radvansky, 1998),[38] as well as an intermediate, gist-like level of a *text-base* (Kintsch & Van Dijk, 1978), in addition to the surface structure representing the precise linguistic form. A situation model represents the accumulated representation of the discourse, potentially including background knowledge and additional inferences, perhaps along central dimension such as time, space, and causation (among others). Crucially for our purposes, however, the situation model does not reflect any structural aspects of the original linguistic form, but merely represents the information expressed. At least in this regard, it is not unlike the Stalnakerian notion of context utilized by Dynamic Semantics, which consists of a set of propositions, construed as possible worlds.

Given our discussion in this paper, this naturally leads us to consider the question of whether, from such a perspective, presuppositions are evaluated at the level of the situation model or the text-base. The former option essentially leads to a view in line with the Global vs. Local Hypothesis above, as the update of situation models with information from a new sentence takes place in a two-step fashion: first, the current sentence is used to construct the 'current model', which is then added to the 'integrated model' that represents prior discourse. To the extent that the Global vs. Local Hypothesis cannot fully account for our data (certainly not on its own), our findings speak against this possibility.

Alternatively, we could assume that presuppositions are evaluated at the level of the text-base. This intermediate level preserves some of the major structural aspects of the linguistic surface form. While propositions are represented at this level in a format that is more abstract than the surface form, crucial structural aspects are still encoded, e.g., whether pairs of expressions were part of the same proposition expressed by a given clause. To the extent that our findings indicate effects based on structural complexity - based on the level of embedding -, it seems quite natural to consider the possibility that presupposition evaluation takes place at such an intermediate level. In fact, it seems promising to see a DRT-style account as providing a characterization of (some aspects of the) structure present in the text-base: connectives and clausal operators express crucial information relating propositions to one another, and since the text-base is assumed to preserve structural information about the propositions introduced by individual clauses, such an extension would seem both welcome and necessary. In other words, the assumption that DRT-style structural representations are cognitively real could naturally be paired with the notion of a text-base from psychological models of discourse processing, in that the latter provides a suitable place in the model for incorporating the insights on presuppositions from formal semantics.[39] While we are not in a position to spell out an

[38] Or related notions such as a mental model (Johnson-Laird, 1983).

[39] An interesting variation of this line of thought, suggested to us by Benjamin Spector (p.c.), would be to consider the relevant structural representations to only be part of the processing model, and have a separate formal account of presuppositions and projection that interacts with them.

integrated theory of a text-base with DRT-style discourse representations, we think that pursuing such an integration seems worthwhile and timely.

The more general upshot of our present discussion is that our results point to the cognitive reality of a structured representation of discourse that is relevant for the processing of presupposed content. The natural level of incorporating such structured information in a comprehensive theory of discourse processing would seem to be one akin to the text-base. We'd like to stress in this context that the proposal here is not to reduce insights from presupposition theory to existing notions in discourse processing models. There are robust and intricate empirical facts about projection that so far have only been successfully analyzed in theories from formal semantics, such as the ones considered here. These insights will certainly need to be preserved, but they also need to be integrated within a broader perspective of discourse processing. We see it as part of the present contribution to initiate more in-depth efforts of integrating central aspects of formal semantic theory with a theory of language processing in context.

5.3.3 Conclusion

As the experimental exploration of presupposition interpretation in online processing is still in its beginnings, many open questions naturally remain. With respect to the issues discussed here, one particularly interesting point is whether different types of presupposition triggers vary in the way that projection proceeds. In particular, considering the distinction between soft and hard triggers just mentioned, would the former display the same types of effects as we found for *again*? The property of triggers being anaphoric seems equally relevant. Standard DRT accounts see all triggers as anaphoric, but there seem to be good grounds for seeing a difference between triggers in this regard (e.g., comparing *again* to *stop* or *realize*), though the diagnosis of such differences is by no means straightforward and has to be considered carefully for each individual case (Heim, 1990, for example, discusses potential evidence for anaphoric aspects of *stop*). Furthermore, given the parallels that DRT posits between pronouns and presuppositions, the question arises of how to relate the present findings to the extensive literature on pronoun interpretation in processing. Future work will need to assess these issues in further detail. As we've tried to stress throughout, an integration of the relevant formal semantic theories with more comprehensive models of discourse processing is generally called for. The contrasting processing hypotheses based on DRT and Dynamic Semantics have provided a very precise illustration of the issues arising from the present experimental results. But we hasten to note that they do not provide final evidence for DRT specifically. First, as noted in section 2.3.1, it requires one particular view of projection in DRT, based on a step-by-step procedure, rather than a parallel consideration of all possible presupposition resolution sites. Furthermore, the experimental results here ultimately will only form part of an overall comparison between the various theories. Some aspects of the predictions of the way projection is handled in DRT,

for example, have been controversial in the literature (see, e.g. Beaver, 2001; von Fintel, 1994; Schlenker, 2011). Potential alternative explanations for the patterns in our data therefore need to be evaluated further as well, as do possible combinations of theoretical frameworks and processing factors.

References

Abrusán, M. (2011). Predicting the presuppositions of soft triggers. *Linguistics and Philosophy*, *34*(6), 491–535.

Abrusán, M., & Szendrői, K. (2013). Experimenting with the king of France: Topics, verifiability, and definite descriptions. *Semantics & Pragmatics*, *6*(10), 1-43.

Abusch, D. (2010). Presupposition triggering from alternatives. *Journal of Semantics*, *27*(1), 37 –80.

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

Bates, D. M. (2005). Fitting linear mixed models in r. *R News*, *5*, 27-30.

Beaver, D. (2001). *Presupposition and assertion in dynamic semantics*. Stanford, CA: CSLI Publications.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, *51*(3), 437–457.

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes*, *28*(4), 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*(3), 434–463.

Chemla, E., & Bott, L. (2013). Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and Cognitive Processes*, *38*(3), 241-260.

Chemla, E., & Schlenker, P. (2012). Incremental vs. symmetric accounts of presupposition projection: an experimental approach. *Natural Language Semantics*, *20*(2), 177–226.

Chierchia, G. (1995). *Dynamics of meaning. anaphora, presupposition and the theory of grammar*. Chicago and London: University of Chicago Press.

Crain, S., & Steedman, M. (1985). On not being led up the garden path. the use of context by the psychological parser. In D. R. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (p. 320-358). Cambridge: Cambridge University Press.

Domaneschi, F., Carrea, E., Penco, C., & Greco, A. (2013). The cognitive load of presupposition triggers: mandatory and optional repairs in presupposition

failure. *Language and Cognitive Processes*, *0*(0), 1-11. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/01690965.2013.830185 doi: 10.1080/01690965.2013.830185

Fox, D. (2008). Two short notes on schlenker's theory of presupposition projection. *Theoretical Linguistics*, *34*(3), 237-252.

Fox, D. (2012). *Presupposition projection from quantificational sentences: trivalence, local accommodation, and presupposition strengthening.* (Ms. MIT & HUJI)

Geurts, B. (1995). *Presupposing* (Ph.D. thesis). University of Stuttgart.

Geurts, B. (1999). *Presuppositions and pronouns.* Amsterdam; New York: Elsevier.

Grice, H. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Speech acts.* New York: Academic 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*(1), 42–55.

Groenendijk, J., & Stokhof, M. (1990). Dynamic montague grammar. In L. Kálmán & L. Pólos (Eds.), *Papers from the second symposium on logic and language* (p. 3-48). Budapest: Akadémiai Kiadó.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases* (Doctoral dissertation, UMass). Retrieved from http://scholarworks.umass.edu/dissertations/AAI8229562

Heim, I. (1983). On the projection problem for presuppositions. *Proceedings of the Second West Coast Conference on Formal Linguistics*, 114-125.

Heim, I. (Ed.). (1990). *Presupposition projection.* University of Nijmegen.

Heim, I. (2006). *Notes on the proviso problem.* (Handout from Milan Meeting, Gargnano.)

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive psychology*, *58*(3), 376–415.

Huang, Y. T., & 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*(8), 1161–1172.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language: Proceedings of the third amsterdam colloquium* (Vol. I, p. 227-321). Amsterdam: Mathematical Center.

Kamp, H., & Reyle, U. (1993). *From discourse to logic.* Dordrecht: Kluwer Academic Publishers.

Karttunen, L. (1973). Presuppositions of compound sentences. *Linguistic Inquiry*, *4*(2), 169-193.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363-394.

Lewis, D. (1979). Scorekeepingin a language game. *Semantics from Different Points of View*.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological*, *124*, 372-422.

Romoli, J., Khan, M., Snedeker, J., & Sudo, Y. (2015). Resolving temporary referential ambiguity using presupposed content. In F. Schwarz (Ed.), *Experimental perspectives on presuppositions* (pp. 67–88). Springer International Publishing.

Romoli, J., & Schwarz, F. (2015). An experimental comparison between presuppositions and indirect scalar implicatures. In F. Schwarz (Ed.), *Experimental perspectives on presuppositions* (pp. 215–240). Springer International Publishing.

Romoli, J., Sudo, Y., & Snedeker, J. (2011). An experimental investigation of presupposition projection in conditional sentences. *Proceedings of SALT*, *21*, 592–608.

Schlenker, P. (2008a). Be articulate: A pragmatic theory of presupposition projection. *Theoretical Linguistics*, *34*(3), 157–212.

Schlenker, P. (2008b). Presupposition projection: The new debate. In T. F. 2008 & S. Ito (Eds.), *Proceedings of SALT 18.* Ithaca, NY: CLC Publications.

Schlenker, P. (2009). Local contexts. *Semantics & Pragmatics*, *3*, 1-78.

Schlenker, P. (2011). The proviso problem: a note. *Natural Language Semantics*, *19*(4), 395-422.

Schwarz, F. (2007). Processing presupposed content. *Journal of Semantics*, *24*(4), 373.

Schwarz, F. (2014). Presuppositions are fast, whether hard or soft - evidence from the visual world paradigm. In T. Snider, S. D'Antonio, & M. Weigand (Eds.), *Semantics and linguistic theory (SALT)* (Vol. 24, pp. 1–22). LSA and CLC Publications. Talk presented at SALT 24.

Schwarz, F. (Ed.). (2015a). *Experimental perspectives on presuppositions.* Springer International Publishing.

Schwarz, F. (2015b). Presuppositions vs. asserted content in online processing. In F. Schwarz (Ed.), *Experimental perspectives on presuppositions* (pp. 89–108). Springer International Publishing.

Schwarz, F., Romoli, J., & Bill, C. (2014). *Scalar implicature processing: Slowly accepting the truth (literally).* Talk presented at SuB 19.

Simons, M. (2001). Simons, m.: On the conversational basis of some presuppositions. In *Proceedings of semantics and linguistics theory 11.*

Singh, R. (2007). Formal alternatives as a solution to the proviso problem. In T. Friedman & M. Gibson (Eds.), *Proceedings of SALT 17* (p. 264-281). Ithaca, NY: Cornell University.

Stalnaker, R. (1973). Presuppositions. *Journal of Philosophical Logic*, *2*(4), 447-457.

Stalnaker, R. (1974). Pragmatic presuppositions. In M. K. Milton & P. K. Unger (Eds.), *Semantics and philosophy.* New York: New York University Press.

Stalnaker, R. (1978). Assertion. In P. Cole (Ed.), *Pragmatics* (Vol. 9, p. 315-322). New York: Academic Press.

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

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Tiemann, S. (2014). *The processing of wieder ('again') and other presupposition triggers* (Doctoral dissertation, University of Tuebingen). Retrieved from `http://hdl.handle.net/10900/57169`

Tiemann, S., Schmid, M., Bade, N., Rolke, B., Hertrich, I., Ackermann, H., ... 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* (p. 581-597).

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

van der Sandt, R., & Geurts, B. (1991). Presupposition, anaphora, and lexical content. In *Text understanding in lilog* (p. 259-296).

von Fintel, K. (1994). *Restrictions on quantifier domains* (PhD thesis). University of Massachusetts Amherst.

Zwaan, R., & Radvansky, G. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162-185.