TEL AVIV UNIVERSITY אוניברסיטת תל אביב

Sagol School of Neuroscience

Prediction during language processing: Beyond activation

Thesis submitted for the degree of

"Doctor of Philosophy"

By

Tal Ness

The thesis was carried out under the supervision of

Professor Aya Meltzer-Asscher

Submitted to the Senate of Tel-Aviv University

July, 2021

Abstract

Prediction during language processing has been extensively studied over the past decades, with a growing body of research focusing on the mechanisms involved in prediction, their properties, and their behavioral and neural manifestation. The current work focuses on a suggested distinction between two qualitatively different lexical prediction processes: "Pre-activation", i.e. activation of lexical/semantic knowledge stored in long-term memory, and "pre-updating", i.e. updating of the sentence's representation built in working memory (WM) to include the predicted content.

In this work I develop and test a model of the joint workings of pre-activation and pre-updating within the routine processing stages of a word in sentence context. According to this model, multiple lexical predictions are simultaneously pre-activated in a graded manner. Pre-updating is only initiated if a certain prediction is highly activated. Namely, if the activation level of a certain prediction passes a retrieval threshold, pre-updating occurs, and the highly activated word is integrated into the sentence representation already when the prediction is generated, rather than when the word appears in the input. If a pre-updated prediction is disconfirmed, prediction failure costs will be incurred (attributed to the need to inhibit the falsely predicted word). Importantly, the threshold for pre-updating is variable; it can differ between individuals (due to factors such as WM abilities), and be adjusted to different situations (due to factors such as predictive validity, noise levels in the input, and task demands), thus controlling the tendency to perform pre-updating in order to balance the benefits of successful predictions and the costs of unsuccessful predictions.

This work is composed of three published papers, reporting a series of behavioral and event-related potentials (ERP) experiments aimed to provide support for the main aspects of the view outlined above. In the first paper, we provide electrophysiological evidence for pre-updating, manifested as an increased P600 amplitude in high (relative to low) constraint sentences, on a verb prior to the highly probable word. We interpret this effect to indicate integration of the highly probable prediction prior to its realization in the input. We additionally show that this P600 effect is positively correlated with participants' reading span scores, suggesting that the tendency to pre-update varies between individuals depending on WM abilities.

In the second paper we examine production onsets in a speeded cloze task (i.e. a sentence completion task in which participants are instructed to produce a completion out loud as quickly as possible). We show that production onsets of the modal response to a sentence (i.e. the most probable completion) are influenced by the strength (cloze probability) and relatedness of a not-produced competitor, the second most probable completion. These results support the idea that multiple predictions are simultaneously pre-activated, and show that the activation level of a predicted word is influenced by the alternative predictions. We additionally provide a computational model to account for production onsets in the speeded cloze task, by adapting and extending Chen and Mirman's (2012) interactive activation and competition (IAC) model.

In the third paper we explore the circumstances under which pre-updating occurs, by employing a speeded cloze task while measuring ERPs on a verb prior to the cloze response production. This allows us to analyze the ERPs based on the specific response produced by the participant in each trial, reflecting

the participant's strongest prediction in that moment. We replicated the increased P600 amplitude in high (relative to low) constraint sentences at the verb leading to the highly probable prediction. Importantly, this pre-updating P600 effect was observed in high constraint sentences (relative to low constraint) even when the participant's strongest prediction in that moment (their produced response) was a low cloze word. These results support a noisy activation race towards a threshold as the mechanism for initiation of pre-updating.

Taken together, these results advance our understanding of the specific mechanisms that underlie prediction during language processing, highlighting the notion that prediction is not one uniform process, and promoting a more nuanced view of prediction.

Contents

1 Introduction

Prediction of upcoming input is a core processing strategy in the human brain, shared across many (if not all) cognitive domains, including perception, sensory-motor processing, and learning (for reviews see Clark, 2015; Hohwy, 2018). The idea that prediction is involved in cognitive processing dates back to the 19th century (James, 1890). In the broadest sense, predictive processing refers to any processing that incorporates information from previous and present input, with inference about future input. In order for this to happen, processing has to be driven not exclusively by the (bottom-up) input, but also by higher-level representations, i.e. accumulated knowledge that can enable the formation of hypotheses about likely upcoming input given previous input, based on experience (see e.g. Bubic, von Cramon, & Schubotz, 2010, for discussion of various definitions of prediction in cognitive science and neuroscience).

Over the past decades, numerous studies have suggested an important role for prediction also in language processing. These studies indicate that while reading or listening to linguistic stimuli, we do not passively wait for the input and process it as it comes, but rather constantly engage in anticipatory processing (for reviews see Huettig, 2015; Kuperberg & Jaeger, 2016; Van Petten & Luka, 2012). Prediction in language processing can take place at various linguistic levels or domains. For example, studies have demonstrated prediction of syntactic structure (e.g. Arai & Keller, 2013; Farmer, Christiansen, & Monaghan, 2006; Garnsey et al., 1997; Gibson & Wu, 2013; Hare et al., 2003; Rohde, Levy, & Kehler, 2011; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Wilson & Garnsey, 2009), semantic content (Altmann & Kamide, 1999, 2007; Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Federmeier & Kutas, 1999; Kamide, Altmann, & Haywood, 2003; Kuperberg, Paczynski, & Ditman, 2011; Matsuki et al., 2011; Metusalem et al., 2012; Paczynski & Kuperberg, 2011, 2012; Xiang & Kuperberg, 2015), phonological information (Allopenna, Magnuson, & Tanenhaus, 1998; DeLong et al., 2005), and orthographic information (DeLong et al., 2005; Dikker, Rabagliati, Farmer, & Pylkkänen, 2010). The current work mostly focuses on what is commonly referred to as lexical predictions, i.e. prediction of the lexical-semantic content of an upcoming word in a sentence, although some of the ideas discussed here may also be relevant for prediction at other levels of representation.

1.1 Evidence of prediction during language processing

An early finding in support of lexical prediction during sentence processing is the decreased reading times observed for highly predictable words relative to unpredictable words (e.g. Ehrlich & Rayner, 1981; Forster, 1981; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985; Stanovich & West, 1983; Traxler & Foss, 2000). The predictability of a word in a sentence context is commonly operationalized using cloze probability values. To obtain these values, participants are given a sentence completion questionnaire termed a ‘cloze questionnaire’, in which the beginning of a sentence is presented, and are asked to provide the first completion that comes to mind. The cloze probability of a word is defined as the percentage of participants who provided it as the sentence’s completion, and this value is considered to reflect how predictable the word is, following the sentence context. An additional measure that is calculated using the cloze task is ‘sentence constraint’, which is defined as the percentage of participants who provided the most common completion for the sentence. This value is considered to reflect the extent to which the sentence context encourages a strong prediction. For example, in sentence

(I) below, the word ‘popcorn’ has a 75% cloze probability since it is produced by 75% of the participants in a cloze task, and the word ‘candy’ has a 10% cloze probability. The sentence constraint for sentence context (I) is 75%, since this is the cloze probability of its most probable completion. This sentence is thus highly constraining, i.e., it encourages a strong prediction. On the other hand, for sentence (II) there is no particularly highly probable completion. The most probable completion for this sentence is ‘book’ and its cloze probability is 25%. Therefore, sentence (II) has a low constraint (25%), which reflects the fact that the sentence does not encourage a strong prediction.[1]

I. Before the movie even started, the kids started to eat the ___
   A. popcorn (75%)
   B. candy (10%)
II. In the classroom, Amy opened the cabinet to take out the ___
   A. book (25%)
   B. notebook (10%)

The studies mentioned above (Ehrlich & Rayner, 1981; Forster, 1981; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985; Stanovich & West, 1983; Traxler & Foss, 2000) found decreased reading times for words with high cloze probability, relative to low cloze probability words. This finding can be explained if probable words are predicted in advance, which means that some of their processing is done prior to their realization in the input, leading to facilitation once they appear (see below for a more specific account of what processes may be carried out predictively).

An additional classical finding in support of prediction comes from the ‘visual world’ eye-tracking paradigm. Altmann and Kamide (1999) showed that when listening to sentences such as ‘the boy will <u>eat</u> the cake’, in which the verb has a strong selectional restriction (i.e. requires an edible object), while looking at a visual array, participants look at the only picture of an edible object in the display (a cake) well before the word ‘cake’ appears in the sentence. This pattern contrasts with the pattern observed for sentences with a less restrictive verb such as ‘the boy will <u>move</u> the cake’. The early looks to the predicted object in the highly constraining contexts indicate anticipation of the upcoming noun (see also Boland, 2005; Kamide, Altmann, & Haywood, 2003).

Finally, a large body of research on prediction in language comes from event-related potential (ERP) experiments. The amplitude of the N400 component was shown by dozens of studies to be decreased for predictable relative to unpredictable words, indicating decreased processing difficulty (see review in Kutas & Federmeier, 2011). In section 1.4.1 below I elaborate further on the N400 component and its relevance to prediction.

1.2 Prediction failure costs and specific word prediction

Importantly, as often noted in the literature, while the findings mentioned above can be attributed to prediction of specific words, they are also compatible with the formation of more general expectations (e.g. Federmeier & Kutas, 1999; Luke & Christianson, 2016; Van Petten & Luka, 2012). Namely, instead

[1] This example was constructed for explanatory purposes, and the presented cloze probabilities are rough estimates.

of leading to prediction of upcoming *words*, a context can lead to activation of some semantic features, for example, the verb "drink" would lead to activation of semantic features such as "liquid" rather than to the prediction of a specific word such as "water". This process would still result in the effects discussed above, since the processing of e.g. "water" will be facilitated due the activation of its feature "liquid" (even if the word 'water' was not directly predicted).

Forming specific word predictions was claimed by some authors to be an implausible processing strategy, since the more specific a prediction is, the more likely it is to be contradicted by the input. Hence, forming specific predictions would have a low 'pay-off' (Forster, 1981; Jackendoff, 2002). Underlying this 'low pay-off' argument is an implicit assumption that the generation of specific predictions consumes some resources, or that the disconfirmation of a prediction incurs prediction failure costs (or both), making the scenario of having a wrong prediction worst then the scenario of not generating a prediction at all (Van Petten & Luka, 2012). Without this assumption, namely if correct predictions are beneficial but incorrect predictions incur no costs, the generation of predictions would be rational regardless of their likelihood of success.

Over the years, several researchers have tried to uncover prediction failure costs. As observed by Van Petten and Luka (2012), in order to make a case for prediction failure costs, it is not sufficient to observe the prevalent finding that the processing of unpredictable words is more demanding than the processing of predictable words (discussed above), since this finding can be explained merely by applying to facilitation for successful predictions. Thus, prediction failure costs can only be evidenced by differences in the processing of words which are unpredictable to the same degree, when a strong prediction could have been formed (i.e. in a high constraint context) relative to when no highly probable prediction was available in the first place (i.e. in low constraint context). In this case, since the comparison is between similarly unpredictable words, there is no facilitation due to predictability, and a difference, if observed, can only be attributed to the existence of the falsely predicted word in the high constraint context.

A few studies revealed little to no evidence of such differences, mostly in behavioral measures (e.g. Frisson, Harvey, & Staub, 2017; Luke & Christianson, 2016, Van Petten et al., 1999). However, ERP studies have produced a robust and well-replicated indication of increased processing demands that are observed when participants are presented with an unexpected word in high relative to low constraint contexts, reflected as a frontal post-N400 positivity (f-PNP; e.g. Federmeier et al., 2007; Kuperberg, Brothers, & Wlotko, 2020). For example, Federmeier and colleagues (2007) compared sentences such as "The children went outside to **play/look** …", to sentences such as "Joy was too frightened to **move/look** …", examining the ERPs elicited by 'look' in both sentences. In the former sentence, 'look' is presented in a high constraint context, since the word 'play' is highly probable (85% cloze probability). In the latter sentence, on the other hand, 'look' is presented in a low constraint context, since the most probable word, 'moved', has only 35% cloze probability. Crucially, although the two sentences differ in constraint (85% versus 35%), the cloze probability of the word 'look' is similar in both (3%). The results showed that while the N400 amplitude elicited by 'look' did not differ between these sentences, a f-PNP was elicited by 'look' in high relative to low constraint context. Since 'look' is similarly unpredictable in both sentences, this result cannot be attributed to the unpredictability of the presented word. Instead, it is attributed to the disconfirmation of a strong prediction in the high constraint context.

The f-PNP findings thus provide evidence of prediction failure costs, i.e. when confronted with an unpredictable word, additional processes need to be recruited in order to overcome a disconfirmed prediction, when there was one. More specifically, in Ness and Meltzer-Asscher (2018a, not included in this dissertation), we have provided evidence suggesting that when a strong prediction fails, inhibition of the disconfirmed prediction is needed in order to enable integration of the actual input, and that this inhibition is reflected in the f-PNP component. Using the cross-modal lexical priming (CMLP) paradigm, we showed that a highly predictable word is strongly activated prior to its anticipated appearance, but is then inhibited if a (congruent) unexpected word appears instead (but see Federmeier & Rommers, 2018, for indication that disconfirmed predictions may not be fully inhibited). Moreover, the inhibition behaviorally measured in the CMLP task was correlated with the amplitude of the f-PNP component in an ERP study with the same materials, in line with the suggestion that the f-PNP reflects this inhibition process. In a similar vein, Kuperberg, Brothers, and Wlotko (2020) have also suggested that the f-PNP component reflects suppression of a disconfirmed prediction. However, while in Ness and Meltzer-Asscher (2018a) we focused on inhibition at the word level, Kuperberg Brothers, and Wlotko (2020) argue for suppression at a higher level of representation, i.e. the event or the situation model.

However, despite these potential costs of prediction failure, evidence suggests that comprehenders do form highly specific predictions, at least under certain circumstances. Namely, comprehenders predict the exact word that is expected to appear, including its phonological form, grammatical features (e.g. gender), etc. (e.g. Delong, Urbach, & Kutas, 2005; Martin et al., 2013; Nieuwland et al., 2018; Nicenboim, Vasishth, & Rösler, 2020; Szewczyk & Wodniecka, 2020; van Berkum et al., 2005; Wicha, Moreno, & Kutas, 2004). For example, Wicha and colleagues (2004) examined ERPs elicited when Spanish native speakers read a determiner (el/la, un/una, las/los), which appears prior to the noun and has to agree with the noun's grammatical gender. Their results show that in sentences that lead to a highly probable noun, determiners with a gender feature that does not match the predictable noun elicit enhanced positivity. These results indicate that the predictions generated were beyond the conceptual level, such that a specific noun was predicted, including its grammatical features. Another study tested ERP responses to words that are form-related (e.g. book-hook) or semantically related (e.g. book-page) to the most predictable word in moderately and highly constraining sentences (Ito, Corley, Pickering, Martin, & Nieuwland, 2016). When the sentences were presented at a rate of 500ms per word, a reduced N400 (relative to unrelated words) was found only for semantically related words, not for form-related ones. A longer SOA (700ms), however, led to a reduced N400 for form-related words as well, but only in highly constraining sentences. These results indicate that predictions of word form, although perhaps requiring more time than semantic predictions and depending on the sentence's constraint, do occur.

1.3 Pre-activation and pre-updating

In recent years, a distinction was suggested to exist between two qualitatively distinct prediction processes, pre-activation and pre-updating (Lau, Holcomb, & Kuperberg, 2013; Kuperberg & Jaeger, 2016), as explained below. I suggest that this distinction can potentially explain how prediction can be a beneficial processing strategy, despite the apparent contradiction between the formation of highly specific predictions (which would have a high probability of failure), and the existence of prediction failure costs (which would make the generation of specific predictions a costly strategy).

"Pre-activation" refers to an increase in the activation level of knowledge stored in long-term memory, i.e. the concept's representation in the lexicon, due to spreading activation as well as more controlled prediction processes. Multiple words can be simultaneously pre-activated. The findings discussed above, demonstrating decreased processing demands for predictable words, are commonly attributed to pre-activation, i.e., the processing of a predictable word is facilitated since the word is already activated to some extent, prior to its bottom-up activation (e.g. Federmeier, 2007; Federmeier & Kutas, 1999; Kuperberg & Jaeger, 2016).

In contrast, the term "pre-updating" is used to refer to the updating of the sentence's representation built in working memory (WM), to include the predicted content. This process was initially posited by Lau Holcomb, and Kuperberg (2013), based on the finding that the N400 effect on target words preceded by prime words was affected by predictive validity (manipulated by changing the proportion of related word pairs in the experimental context). Specifically, their results showed that the reduction in N400 amplitude for related prime-target pairs was greater in an experimental block in which the proportion of related pairs was high, thus encouraging prediction, relative to a block with a lower proportion of related pairs (making prediction less beneficial). The authors suggested that these results indicate pre-updating, under the assumption that high predictive validity leads to the pre-updating of WM representations, and the updated WM representations in turn affect activations, leading to a greater reduction in N400 amplitude for related pairs. However, there is no reason to assume that top-down control cannot directly enhance or limit the spread of activations, without the mediation of WM representations (see e.g. Van Berkum, 2009). Therefore, these results do not provide direct evidence of pre-updating. In this work, my first goal was thus to provide evidence of pre-updating, by more directly looking at integration processes rather than activations (Ness & Meltzer-Asscher, 2018b, below).

A key difference between pre-activation and pre-updating is that while the former results in priming of multiple words, the latter entails commitment to a specific prediction. Accordingly, I propose that only a pre-updated prediction incurs processing costs if disconfirmed. If no commitment is made about an upcoming word, then processing difficulty associated with an unexpected word would only depend on that word's activation level (correlated by hypothesis with its cloze probability): less probable words need to accumulate more activation in order to be retrieved. However, there would be no additional costs related to any alternative prediction that was generated. In contrast, if commitment was made, i.e. if a prediction was pre-updated into the sentence representation, then the unexpected word that appears in the input cannot be seamlessly processed. Inhibition of the pre-updated prediction is required in order to enable integration of another word instead (see further discussion in Ness & Meltzer-Asscher, 2018a). Thus, unlike when only pre-activation has taken place, the disconfirmation of a pre-updated prediction does incur additional processing costs.

The proposed view of pre-activation and pre-updating is summarized in Figure 1. I view pre-activation as an "unavoidable" or "automatic" process that occurs whenever linguistic input is processed. However, only when activation of some predicted content reaches a certain threshold, this content is updated into the sentence representation. In many cases, the threshold is reached only with the aid of bottom-up input, namely when the predicted word is encountered in the sentence. However, the retrieval threshold can also be reached prior to the realization of the word in the input, when the sentence being processed is highly constraining (i.e. leading to a stronger prediction).

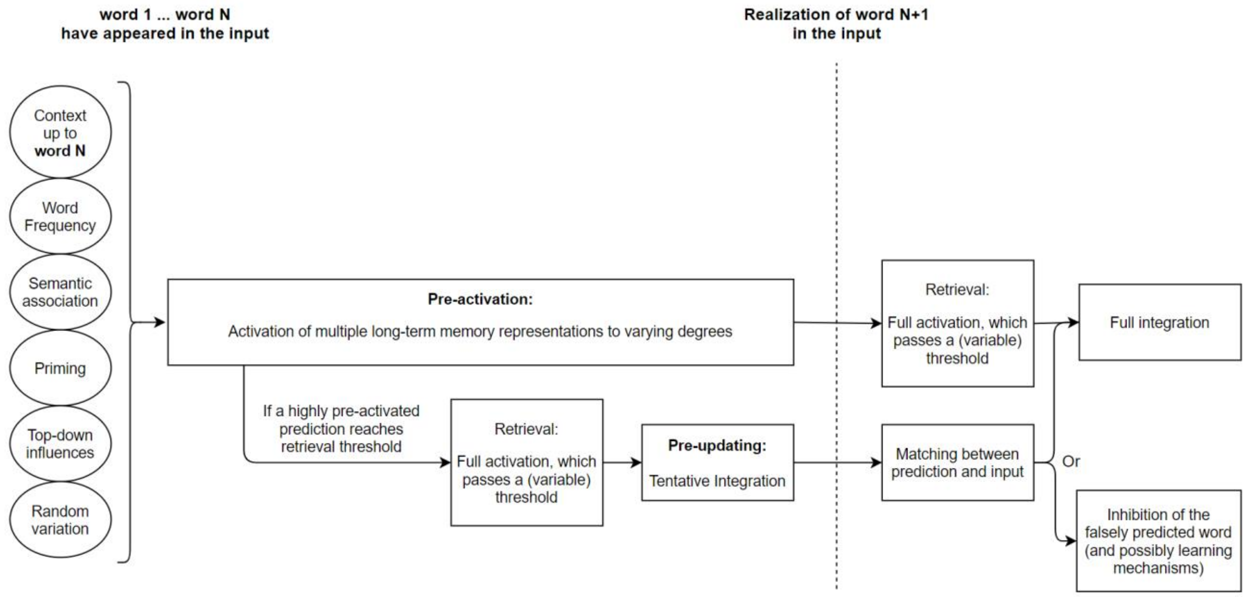

Figure 1. An outline of prediction processes within the processing stages of a word.

I additionally propose that the retrieval threshold which initiates pre-updating can potentially vary between individuals, such that for a similar prediction strength, different comprehenders would be more or less likely to pre-update, depending on their individual traits. For example, comprehenders with greater WM abilities and better ability to handle prediction failure costs, if these would be incurred, may be more inclined to pre-update, namely have a lower retrieval threshold. This suggestion accords with proposals of individual differences in threshold-based mechanisms in other cognitive domains, such as decision making and response speed-accuracy tradeoff (e.g. Heitz, 2014; Jackson et al., 2016).

The retrieval threshold may also vary between different situations, depending on factors such as task demands, predictive validity and noise levels, which influence how beneficial it is to engage in strong prediction in a given situation. For example, in a situation where predictive validity is low (i.e. when strong predictions are often violated, namely, there is a high proportion of high constraint contexts that end with an unexpected word), it may not be beneficial to commit to a specific prediction even in high constraint contexts, and therefore the threshold for pre-updating may be raised. Several studies have shown that predictions can be adapted to different situations (e.g. Brothers et al., 2019; Brothers, Swaab, & Traxler, 2017; Hutchison 2007; Lau, Holcomb, & Kuperberg, 2013; Neely, 1977; Schwanenflugel & Shoben, 1985). However, the manipulations in these studies did not specifically target pre-updating or prediction failure costs. In a recent study (Ness & Meltzer-Asscher 2021c, not included in this dissertation), we have shown that prediction failure costs decrease when the participant estimates that the predictive validity in the experimental context is low. These results provide initial indication that comprehenders may alleviate prediction failure costs when prediction validity is low by raising the threshold for pre-updating, thus avoiding pre-updating and preventing the need to perform inhibition when the prediction is disconfirmed.

Thus, in situations where it is not likely to be beneficial to form a specific prediction and commit to it, due to a high probability of failure (e.g. in a low constraint context, and/or when predictive validity is low), prediction will only manifest in graded pre-activation, and will not incur failure costs (even if the most probable word does not appear). Only when committing to a specific prediction has a high probability of success (e.g. in a high constraint context, when predictive validity is not low), such commitment is engaged in, i.e. pre-updating occurs. In these scenarios, the benefits of successful prediction are more frequent than the costs of failure.

I also propose that the distinction between pre-activation and pre-updating can provide an explanation for why prediction failure costs where not always observed in the literature (e.g. Frisson, Harvey, & Staub, 2017; Luke & Christianson, 2016; , Van Petten et al., 1999). First, under the current view, prediction failure costs are not gradual throughout the entire scale of sentence constraint; they only occur at the higher end of the scale. For example, consider an unpredictable word with e.g. 5% cloze probability, appearing in a moderately constraining sentence with e.g. 50% sentence constraint, compared to a similarly low cloze probability word (5%) appearing in a highly constraining sentence with e.g. 85% sentence constraint. In both cases there is some prediction error, but this does not entail that prediction failure costs will be proportional to the magnitude of the prediction error in each sentence. Instead, if pre-updating only occurs for very strong predictions (e.g. predictions with cloze probability above 70%), then in sentences with 50% constraint there would be no prediction failure costs at all. This may hinder detection of prediction failure costs in analyses designed to test for correlation between prediction error (or an equivalent measure) and prediction failure costs, throughout the entire range of constraint, i.e. including sentences with moderately and even low constraint (e.g. Luke & Christianson, 2016). Moreover, the occurrence of prediction failure costs depends on the tendency to pre-update, which may vary substantially depending on the specific details of each experiment (e.g. predictive validity, task demands). Specifically, this may contribute to the lack of evidence for prediction failure costs in experiments employing eye-tracking while reading (Frisson, Harvey, & Staub, 2017; Luke & Christianson, 2016). In such experiments, sentences are presented in their entirety and for unlimited time, and the participant can control their reading rate as well as regress to previously read material. For this reason, participants have no rational incentive to engage in strong prediction. Namely, when the participant controls their intake of the input, there may be no reason to 'run ahead' and integrate a prediction prior to its perception, since it can simply be read at a pace that best corresponds to the participant's processing. Notably, this is in contrast to spoken language, in which the auditory input rapidly disappears, as well as to visual word-by-word presentation at a fixed rate, which is common in psycholinguistic experiments (and specifically in the ERP literature on prediction).

1.4 ERP correlates of activation and integration

Studies 1 and 3 reported below use event-related potentials to investigate prediction processes. I thus briefly discuss here the most relevant findings from the two ERP components I used – the N400 and the P600. For more detailed reviews see Kutas & Federmeier (2011) and Brouwer et al. (2017).

1.4.1 The N400

The N400 ERP component is a centro-parietal negativity peaking between 300 and 500 ms after stimulus onset. This highly studied component was initially found for semantically incongruent sentence-final words (e.g. the word 'socks' in the sentence: "he spread the warm bread with socks") relative to congruent ones (Kutas & Hillyard, 1980). However, it was later shown that the N400 is not specifically elicited by semantic incongruency. Rather, the component is also manifested in response to congruent words, showing a graded amplitude that inversely correlates with the word's cloze probability (Kutas &

Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984, Delong, Urbach & Kutas, 2005; Wlotko & Federmeier, 2012). Moreover, the N400 elicited by mid-sentence content words in an unfolding sentence becomes progressively smaller as the sentence context becomes increasingly constraining (i.e. leading to stronger predictions) (Van Petten & Kutas, 1990, 1991; Van Petten, 1993; Dambacher, Kliegl, Hofmann & Jacobs, 2006). Taken together, these findings suggest that the amplitude of the N400 is inversely correlated with the activation level of a word, thus reflecting the effort exerted to retrieve the word (see e.g. Brouwer et al., 2012, 2017, Delong, Urbach, & Kutas, 2005, Federmeier and Laszlo, 2009; Kutas and Federmeier, 2000, 2011). Namely, the more pre-activated a word is, the easier it is to retrieve, and the lower the N400 amplitude it elicits.

Another robust finding is that unpredictable words that are semantically related to the predictable word elicit a smaller N400 compared to unrelated words. This pattern was demonstrated both for low-cloze congruent words (Thornhill & Van Petten, 2012), and for anomalies (Kutas et al., 1984; Kutas & Hillyard, 1984; Federmeier & Kutas, 1999). For example, in a sentence such as "They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of **palms/pines/tulips**", the word 'pines', elicits decreased N400 amplitude compared to the word 'tulips', due to the greater relatedness of the former to the predictable continuation 'palms' (Federmeier & Kutas, 1999). This result can be explained by assuming overlap of pre-activated features (e.g. 'a tree') between the predicted word and related words (Federmeier & Kutas, 1999), or spreading activations from the predicted word to related words, causing the related word to be activated and therefore easier to retrieve, thus eliciting a smaller N400 relative to an unrelated word.

Notably, the N400 was argued by several authors to reflect integration demands, rather than retrieval costs (e.g. Brown & Hagoort, 1993; Hagoort et al., 2009). However, as noted by Brouwer and colleagues (2012, 2017), this suggestion cannot account for the reduced N400 observed for anomalies when they are related to a predictable word (Kutas et al., 1984; Kutas & Hillyard, 1984; Federmeier & Kutas, 1999). Since these words create anomalous sentences, their integration should be difficult regardless of their relatedness to a predictable word, and they should thus elicit a large N400, contrary to fact. In contrast, as explained above, the retrieval view predicts a small N400 for these words, as their retrieval is facilitated by the pre-activation of the predicted word, which is related to them.

Additionally, the suggestion that the N400 indexes integration rather than retrieval costs is also challenged by findings showing that sentences which are anomalous due to thematic role reversal of a highly probable role assignment, do not elicit increased N400 amplitudes, even though integration should be difficult (e.g. Chow & Phillips, 2013; Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Kuperberg et al., 2007; Van Herten, Kolk, & Chwilla, 2005). For example, Hoeks and colleagues (2004) have shown that in anomalous Dutch sentences such as (III) below, the verb (*geworpen*, 'thrown') does not elicit increased N400 amplitude, relative to congruent control sentences such as (IV) (these sentences instead elicit increased P600 amplitudes, see discussion of the "semantic P600" below). In these sentences, the integration of 'thrown' should be more difficult in (III) than in (IV), but this is not reflected in the N400. In contrast, retrieval should be similarly easy in both sentences, since 'thrown' is highly activated by the words 'javelin' and 'athlete' (regardless of the structure in which they appear), in line with the absence of an N400 effect. Note, that the absence of increased N400 in such sentences does not rule out the possibility that the N400 reflects both retrieval and integration, to some extent (see Nieuwland et al., 2020), but it does suggest that the N400 does not chiefly reflect integration.

III. *De speer heeft de atleten geworpen*
the javelin has the athletes thrown
'The javelin threw the athletes'

IV. *De speer werd door de atleten geworpen*
the javelin was by the athletes thrown
'The javelin was thrown by the athletes'

1.4.2 The P600

The P600 ERP component is a positive deflection in the ERP, with a posterior distribution over the scalp, usually observed within a time-window of 500-900 ms post stimulus onset. This component was initially observed in response to syntactic anomalies such as violation of subcategorization constraints (e.g. Osterhout & Holcomb, 1992) and agreement errors (e.g. Hagoort, Brown & Groothusen, 1993), as well as in "Garden path" sentences, where an initial structure needs to be reanalyzed (e.g. Osterhout & Holcomb, 1992, Osterhout, Holcomb, & Swinney, 1994; Hagoort, Brown & Osterhout, 1999). Importantly, however, the P600 was also shown to be elicited in grammatical sentences that do not involve reanalysis. One such example is the increased P600 amplitude observed when a long-distance dependency is completed (e.g. Felser, Clahsen, & Munte, 2003; Fiebach, Schlesewsky & Friederici, 2002; Gouvea, Phillips, Kazanina, & Poeppel, 2010; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005). For instance, an increased P600 amplitude is measured on the verb "imitated" in a sentence such as (V) relative to (VI) (Kaan, Harris, Gibson, & Holcomb, 2000). The difference between these sentences is that at the verb in (V), integration of the filler occurs. In contrast, the processing of the verb in (VI) does not include this additional process. Thus, there are more integration demands at this point in (V) relative to (VI).

V. Emily wondered who the performer in the concert had imitated __ for the audience's amusement.
VI. Emily wondered whether the performer in the concert had imitated a pop star for the audience's amusement.

The P600 was also found in "semantic illusion" contexts, namely in syntactically sound sentences that are semantically anomalous due to thematic role reversal or thematic violations as in example III relative to IV below (the "Semantic P600", see e.g. Chow & Phillips, 2013; Hoeks, Stowe, & Doedens, 2004; Kuperberg et al., 2007). These and other findings have led to the suggestion that the P600 amplitude reflects integration difficulty (Brouwer, Fitz, & Hoeks, 2012; Kaan et al., 2000; For further discussion regarding the functional nature of the P600, also see Chow & Phillips, 2013).

1.5 The current studies

In the current studies (the three published papers included in chapters 2-4 below), I focus on the suggested distinction between pre-activation and pre-updating, aiming to establish an understanding of the mechanisms underlying these processes. Here I very briefly outline the research questions and results of each study.

1.5.1 Predictive pre-updating and working memory capacity: Evidence from event-related potentials (Ness & Meltzer-Asscher, 2018b)

In this paper we report an ERP experiment, aimed to provide evidence of pre-updating. As discussed above, pre-updating involves integration of a strongly predicted word into the sentence representation, prior to the realization of the word in the input. Since the P600 component was suggested to reflect integration processes, we hypothesized that pre-updating would be reflected in increased P600 amplitude, on a word prior to a highly predictable continuation. Thus, we presented participants with high and low constraint sentences (presented word-by-word), and examined the ERPs of both the highly predictable noun in the high constraint context (relative to low constraint), and the verb prior to this noun, where the strong prediction can be generated. As expected, at the noun we observed decreased N400 amplitude in high relative to low constraint contexts, reflecting facilitation due to the predictability of the presented noun in the high constraint context. Crucially, on the preceding verb we observed increased P600 in high relative to low constraint contexts, reflecting, according to our interpretation, additional integration processes prior to the highly probable word in high relative to low constraint sentences, presumably due to integration of the predicted noun. Additionally, we tested the participants' reading span, a measure aimed to reflect WM abilities. The pre-updating P600 effect was correlated with reading span, such that individuals with higher reading span demonstrated a larger P600 effect, indicating that individuals with better WM abilities have a higher tendency to pre-update. These results provide evidence for pre-updating, and for individual differences in the tendency to pre-update.

1.5.2 Love thy neighbor: facilitation and competition between parallel predictions (Ness & Meltzer-Asscher, 2021a)

This paper focused on the mechanisms of pre-activation. As mentioned above, we assume, with others, that multiple predictions are pre-activated in parallel (unlike pre-updating, which is limited to a specific prediction). Thus, the study reported in this paper was aimed to explore the interactions between simultaneously activated predictions. The study employed a speeded cloze task (i.e. a sentence completion task in which participants are instructed to produce a completion out loud as quickly as possible), looking at the production onset of cloze responses. We began with an exploratory analysis, using pre-existing data, in order to develop a specific hypothesis about how alternative predictions affect each other. We then followed up with a replication experiment, for which the analysis (corresponding to the specific hypothesis) was pre-registered. The results show that production onsets of the modal response to a sentence (i.e. the most probable completion) are influenced by the strength (cloze probability) of a not-produced competitor, the second most probable completion. When these two words are highly related, the stronger the competitor is, the more it facilitates the production of the modal response (i.e. earlier production onset). However, when the two words (the modal and the competitor) are relatively unrelated, the stronger the competitor is, the more it delays the production of the modal response (i.e. later production onset). These findings demonstrate that multiple predictions are pre-activated simultaneously, affecting the activation level of each other. We provide an account for this pattern within the interactive activation and competition (IAC) framework, by adapting a computational model previously employed to account for neighborhood effects on single-word recognition and production (Chen & Mirman, 2012). We additionally extend this model to account for previously observed effects in the speeded cloze task (Staub et al., 2015).

1.5.3 From pre-activation to pre-updating: A threshold mechanism for commitment to strong predictions (Ness & Meltzer-Asscher, 2021b)

Capitalizing on the pre-updating P600 effect we demonstrated in the study discussed above (1.5.1), we conducted an ERP study combined with a speeded cloze task, aiming to test the hypothesis that pre-updating is initiated by an activation threshold mechanism. Participants were presented with high and low constraint contexts (presented word by word), followed by a blank line prompting them to produce a completion. We used the specific response produced by the participant in each trial, reflecting the participant's strongest prediction in that moment, to analyze the ERPs on a verb prior to the production prompt. Increased P600 amplitude was observed in high (relative to low) constraint sentences, and this effect was correlated with reading span, replicating our previous results. Importantly, the pre-updating P600 effect was observed in high constraint sentences (relative to low constraint) even when the participant's strongest prediction in that moment (their produced response) was a low cloze word. These results support a noisy activation race towards a retrieval threshold as the mechanism for initiation of pre-updating.

2 Papers

2.1 Predictive pre-updating and working memory capacity: Evidence from event-related potentials

Tal Ness and Aya Meltzer-Asscher (2018b)
Journal of Cognitive Neuroscience, 30(12), 1916-1938

Abstract

It was recently proposed that lexical prediction in sentence context encompasses two qualitatively distinct prediction mechanisms: "pre-activation", namely activating representations stored in long-term memory, and "pre-updating", namely updating the sentence's representation, built online in working memory, to include the predicted content (Lau, Holcomb, & Kuperberg, 2013). The current study sought to find evidence for pre-updating and test the influence of individual differences in working memory (WM) capacity on the tendency to engage in this process. Participants read strongly and weakly constraining sentences. Event-related potentials were measured on the predictable noun as well as on the preceding verb, where the prediction is generated. Increased P600 amplitude was observed at the verb in the strongly constraining sentences, reflecting integration of the predicted upcoming argument, thus providing evidence for pre-updating. This effect was greater for participants with higher WM capacity, indicating that the tendency to engage in pre-updating is highly affected by WM capacity. The opposite effect was observed at the noun, i.e. for participants with higher WM span, a greater decrease in P600 amplitude in the strongly constraining sentences was observed, indicating that the integration of a pre-updated word was easier. We discuss these results in light of previous literature and propose a plausible architecture to account for the interplay between pre-activation and pre-updating, mediating the influence of factors such as WM capacity.

1) Introduction

Over the past couple of decades, studies focusing on prediction processes led to an increasingly strong consensus regarding the role of prediction in sentence processing. It is now widely assumed that in the course of comprehending a sentence, we do not passively wait for the input and process it as it comes, but rather constantly engage in some form of anticipatory processing. A classical finding demonstrating this is the decreased reaction times observed for predictable as compared to unpredictable words in a sentence (e.g. Ehrlich & Rayner, 1981; Forster, 1981; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985; Stanovich & West, 1983; Traxler & Foss, 2000). Similarly, the amplitude of the N400 event-related potentials (ERP) component elicited by a word has been shown to inversely correlate with the word's cloze probability, meaning that N400 amplitude decreases as the word's predictability increases (e.g. Delong, Urbach & Kutas, 2005; Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984; Wlotko & Federmeier, 2012). Interestingly, the amplitude of the N400 is not decreased only for predictable words. Unpredictable words that are semantically related to a predictable word elicit a smaller N400 compared to unrelated words (e.g. Thornhill & Van Petten, 2012), and this is true even for anomalous words (e.g. Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Kutas et al., 1984). These findings indicate that the amplitude of the N400 reflects an architecture that involves pre-activation, i.e. that a decreased N400 likely reflects easier retrieval of words that were already activated, either due to spreading activation from the predicted word to related words or due to activation of shared features/concepts. In addition to "spreading" or shared activation, the activation level of a word, reflected in its N400 amplitude, is also affected by higher-level prediction processes. This is shown by findings of N400 sensitivity to factors that influence top-down control, such as predictive validity (e.g. Lau, Holcomb, & Kuperberg, 2013).

1.1) Pre-activation vs. pre-updating

Although the general notion of prediction is largely accepted in the psycholinguistic literature, the precise nature of the processes involved, as well as the extent and ubiquity of prediction during sentence processing, are still under ongoing debate. Recently, Lau et al. (2013) suggested a distinction between two qualitatively distinct mechanisms of prediction, later referred to as "pre-activation" and "pre-updating" (Kuperberg & Jaeger, 2016). "Pre-activation" refers to an increase in the activation level of knowledge stored in long-term memory, i.e. the concept's representation in the lexicon, due to spreading activation as well as more controlled prediction processes. In contrast, the term "pre-updating" is used to refer to the updating of the sentence's representation built in working memory (WM), to include the predicted content.

Two different accounts were put forward regarding the relation between pre-activation and pre-updating. While Lau et al. (2013) assume that pre-activation leads to pre-updating, Kuperberg and Jaeger (2016) assume the opposite order. To illustrate this, Kuperberg and Jaeger consider the following sentence fragment: 'The day was breezy so the boy went outside to fly a …'. We can hypothesize two possible processing scenarios occurring after encountering this fragment. One possibility, adopted by Lau et al. (2013), is that the partial representation of the event (<boy flies>) leads to pre-activation of the lower-level representation of 'kite'. This pre-activated lexical representation is then pre-updated, i.e. it enters WM to be integrated with the sentence's representation, which in turn affects higher representational levels, including updating of the event representation (to be <boy flies kite>). The other possibility, adopted by Kuperberg and Jaeger (2016), is that the partial representation of the event (<boy flies>) is first pre-updated (to be <boy flies kite>), which then causes pre-activation of the lower-level representation of 'kite'. Distinguishing between these accounts is difficult; although they differ in the assumed order in which the different representational levels are updated, they ultimately lead to very similar predictions. Here, we adopt the view that pre-activation precedes pre-updating, and that pre-updating can then "propagate" to higher representational levels.

An important difference between pre-activation and pre-updating is that while the former entails priming of multiple entities, the latter entails commitment to a specific prediction, which would incur processing costs if disconfirmed. If no commitment is made about an upcoming word, then processing difficulty associated with an unexpected word should only depend on this word's activation level (correlated by hypothesis with its cloze probability). Indeed, several studies have shown that low-cloze words elicit similar N400 amplitudes regardless of whether their preceding context is strongly or weakly constraining. Crucially, however, it was also found that low-cloze words that follow strongly as opposed to weakly constraining contexts elicit a late anterior positivity, the frontal post-N400-negativity (fPNP).

This component was argued to reflect an additional cost of prediction failure, exhibited only when the comprehender committed to a specific prediction (e.g. Brothers, Swaab, Traxler, 2015; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Ness & Meltzer-Asscher, in press. see Van Petten & Luka, 2012 for a review).

Thus, as opposed to pre-activation, which is graded and can occur to different degrees depending on prediction strength and specificity, pre-updating is an "all or nothing" mechanism, which compels commitment. This conjecture stems from the inherent properties of the architectures of long-term memory and working memory, which are vastly different. While an immeasurable number of representations is simultaneously stored in long-term memory, the capacity of WM is highly limited (e.g. "the magical number seven" suggested by Miller, 1956, or "the magic number four", Cowan, 2010; Green, 2017) and therefore it is unlikely that many competing predictions can simultaneously be pre-updated.[2] Based on the properties of these two different memory systems, we view pre-activation as an "unavoidable" process that occurs whenever linguistic input is processed. However, only when activation of some predicted content reaches a certain threshold, this content is pre-updated; at this threshold the predicted content is, for all intents and purposes, retrieved, and is therefore integrated into the sentence's representation in WM. It should be noted that we do not claim that pre-updating is a mechanism designated for prediction per-se. As any sentence is incrementally processed, each word undergoes activation, retrieval and integration. The notion of pre-updating, as we see it, merely means that retrieval can be achieved even without the need to wait for bottom-up activation, when top-down activation is strong enough to reach the retrieval threshold. In this case, the retrieved content will simply move on to the following processing stages, i.e. structure building, semantic integration, thematic role assignment, etc., similarly to a word that had actually appeared in the input (but see Discussion for potential differences). This retrieval threshold will be more likely reached (prior to the realization of the word in the input) when the sentence being processed is more constraining (leading to a stronger prediction) and when predictive validity is high. The threshold could also potentially vary between individuals, such that for similar prediction strength, different comprehenders would be more or less likely to pre-update.

The findings mentioned above, demonstrating priming effects in reaction times and in N400 amplitudes, can be explained by pre-activation alone, without the need to appeal to an additional process of pre-updating. This is also the case for other manifestations of prediction in sentence processing, such as anticipatory eye movement in the visual world paradigm (e.g Altmann & Kamide, 1999; Boland, 2005; Kamide, Altmann, & Haywood, 2003). As all of these findings reflect activations of stored representations, without direct examination of representations built in WM, they can only provide evidence for pre-activation. Lau et al. (2013) presented findings which they claimed were more relevant to pre-updating. The authors showed that the N400 effect in word pairs was affected by predictive validity, manipulated by changing the proportion of related word pairs in the experimental context. Greater facilitation, reflected by a reduced N400 amplitude, was observed in the context which contained a larger proportion of related words, thereby encouraging prediction. This finding was suggested by the authors to indicate pre-updating, under the assumption that high predictive validity leads to pre-updating of WM representations, and the updated representation in turn affects activations, leading to a greater relatedness effect, i.e. reduced N400. However, there is no reason to assume that top-down control cannot

[2] For discussion of a different approach, see Section 4.5.

directly enhance or limit the spread of activations, without the mediation of WM representations (see e.g. Van Berkum, 2009). Therefore, these results do not provide unambiguous evidence in favor of pre-updating. It should also be noted that the processing of word pairs may differ substantially from sentence processing, with participants adopting prediction strategies specific to the task (e.g. lexical or semantic decision to the second word in the pair).

The goal of the current study was therefore to find more direct support for pre-updating by looking for evidence for integration processes prior to the onset of a highly predictable word in a sentence, as well as for decreased integration demands when encountering the predicted word. We hypothesized that these effects would be reflected in the P600 ERP component, which is a measure for integration processes. The P600 component is a positive deflection in the EEG, with a posterior distribution over the scalp. This component was initially observed in response to syntactic anomalies such as violation of subcategorization constraints (e.g. Osterhout & Holcomb, 1992) and agreement errors (e.g. Hagoort, Brown & Groothusen, 1993), as well as in "Garden path" sentences, where an initial structure needs to be reanalyzed (e.g. Osterhout & Holcomb, 1992, Osterhout, Holcomb, & Swinney, 1994; Hagoort, Brown & Osterhout, 1999). Additionally, the P600 was shown to be elicited in grammatical sentences that do not involve reanalysis. Importantly, increased P600 amplitude is observed when a long-distance dependency is completed (e.g. Felser, Clahsen, & Munte, 2003; Fiebach, Schlesewsky & Friederici, 2002; Gouvea, Phillips, Kazanina, & Poeppel, 2010; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005). For example, an increased P600 amplitude is measured on the verb "imitated" in a sentence such as (1a) relative to (1b) (Kaan, Harris, Gibson, & Holcomb, 2000). The difference between these sentences is that at the verb in (1a), integration with the filler occurs. In contrast, the processing of the verb in (1b) does not include this additional process. Thus, there are more integration demands at this point in (1a) relative to (1b).

1. a. Emily wondered <u>who</u> the performer in the concert had **imitated** __ for the audience's amusement.
   b. Emily wondered whether the performer in the concert had **imitated** a pop star for the audience's amusement.

The P600 was also found in "semantic illusion" contexts, namely syntactically sound sentences that are semantically anomalous due to thematic role reversal or thematic violations (the "Semantic P600", see e.g. Chow & Phillips, 2013; Hoeks, Stowe, & Doedens, 2004; Kuperberg et al., 2007). These and other findings have led to the suggestion that P600 amplitude reflects integration difficulty (Brouwer, Fitz, & Hoeks, 2012; Kaan et al., 2000; For further discussion regarding the functional nature of the P600, also see Chow & Phillips, 2013).[3]

In the current study we adopt the suggestion that the P600 reflects some form of integration, without committing to a specific characterization of the processes reflected by the P600. In our case, it is sufficient to assume that any of the processes involved in integrating a word - syntactic structure building, dependency formation, thematic role assignment, semantic integration, etc. - affects P600 amplitude, in order to hypothesize that pre-updating of a predicted word will be reflected by increased

[3] The P600, characterized by a posterior scalp distribution, should be distinguished from the fPNP described above, indexing the costs of failed prediction.

P600 amplitude at the point where pre-updating occurs, followed by decreased P600 amplitude when the already pre-updated word is encountered.

1.2) A reanalysis of EEG data

In order to test the feasibility of finding a measurable P600 difference due to pre-updating, we reanalyzed EEG data collected for a previous study (Ness & Meltzer-Asscher, in press). The data were collected from 24 participants (14 male), native Hebrew speakers, with an average age of 25.7 (range: 19-37). Eighty-four experimental sentences were used, with sentence constraints ranging from 53.6% to 100%. For the new analysis, we divided the sentences such that half were classified as high-constraint and half as low-constraint (example sentences are provided in Appendix A). The average constraint in the high-constraint sentences was 89.5% (range: 80% - 100%), and the average constraint in the low constraint sentences was 68% (range: 50% - 80%). The critical word in the sentences was the verb, on which the prediction was generated. Verbs in the high-constraint and low-constraint conditions were matched on length ($p = .799$), frequency ($p = .898$, corpus: Linzen, 2009) and position in the sentence ($p = .906$, measured in number of words). Due to the design of the original experiment, after the critical verb the experimental sentences continued with either the predicted word, a congruent but unexpected word, or an anomalous word. Sentences were presented word-by-word in the middle of the screen for 200ms, with a 300ms ISI. For a more detailed description of the materials, procedure and EEG recording see Ness and Meltzer-Asscher (in press).

Based on the typical time-window and scalp distribution of the P600, mean amplitudes over 500-800ms from the verb onset for all centro-parietal and parietal electrodes were entered into a repeated-measures ANOVA with the factors Constraint (High, Low) and Electrode (9 levels: CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8). A main effect of Constraint was found ($F_{(1,22)} = 7.05$, $p = .014$), such that verbs in the high-constraint sentences yielded a larger P600 amplitude than those in the low-constraint sentences. Grand averaged ERPs at a representative electrode and scalp distribution of the P600 component are provided in Figure 1.

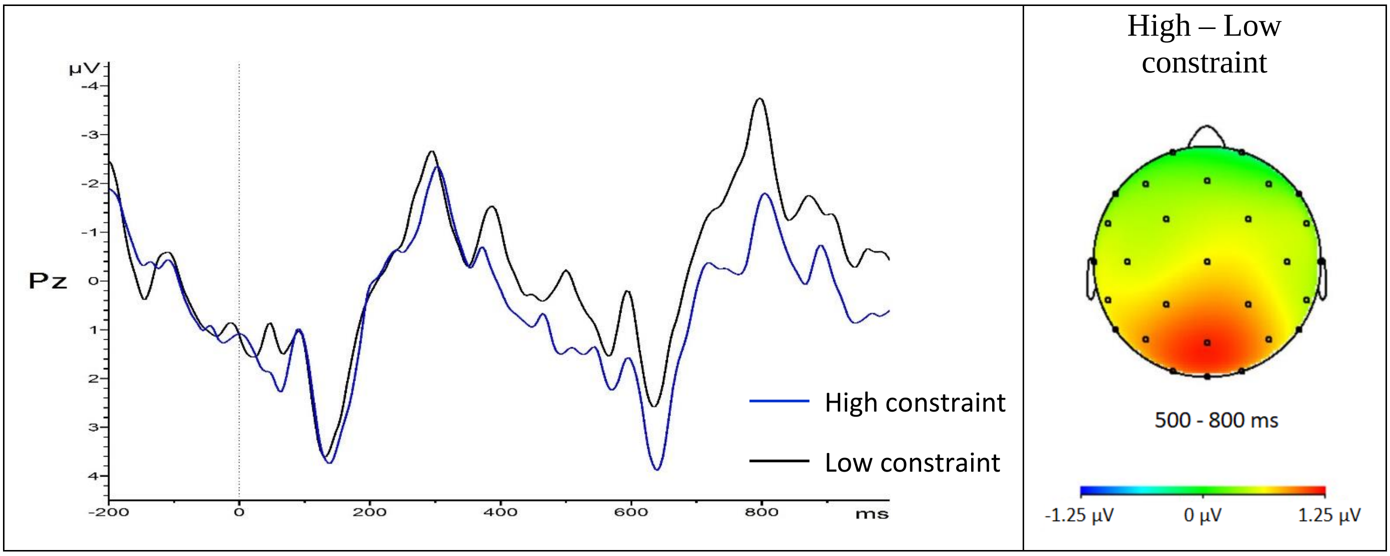

Figure 1: Grand averaged ERPs and scalp distribution for the reanalysis results

These results provide initial evidence for pre-updating, as they indicate that more integration processes take place at the verb when it has a highly predictable argument (high-constraint), relative to when there is no highly predictable argument (low-constraint). A caveat to these results is that the critical word (i.e. the verb), as well as the following words, were not identical between the high- and low-constraint sentences. Additionally, we could not compare the ERPs to the actual predicted word between the high- and low-constraint sentences, to test whether its integration is easier when pre-updating had already occurred at the verb. This was due to the fact that in the original design, the predicted word appeared in only one third of the trials (which would result in fourteen trials in each condition), and due to the fact that the predicted word was presented immediately following the verb and its signal would therefore be contaminated by the P600 effect at the verb.

1.3) The current study

The current study was designed to overcome the limitations of the reanalysis we conducted, as well as replicate the result observed on the verb. Moreover, as pre-updating involves representations built online in WM, we hypothesized that the extent to which an individual tends to engage in pre-updating depends on their WM capacity.

In the present experiment, participants read strongly and weakly constraining sentences in Hebrew. ERPs were measured on the predictable noun phrase (NP) as well as on the preceding verb. These two critical words were separated by an additional word, the Hebrew accusative case marker, in order to make sure that the signal on the NP is not contaminated by effects originating at the verb. Participants' WM capacity was assessed via a reading span test.

If pre-updating indeed occurs, and is more likely to occur when the prediction is stronger, a P600 effect is predicted on the verb in the strongly constraining sentences, reflecting integration of the predicted upcoming argument (the following NP). On the NP, an opposite effect is predicted, namely increased P600 in the low-constraint sentences, reflecting the benefits of pre-updating, i.e. easier integration of a word that had already been pre-updated. Moreover, if participants with higher reading span exhibit a higher tendency to engage in pre-updating, then the effects should be greater for these participants.

2) Methods

2.1) Participants

Participants were 37 Tel-Aviv University students (18 males), all native Hebrew speakers, with an average age of 25.4 (range: 19-40). Participants were given course credit or paid 60 NIS (~$15) for their participation. One participant was excluded from the analysis due to excessive artifacts. The experimental protocol received approval from the Ethics Committee in Tel Aviv University.

2.2) Materials

The materials consisted of 52 sentence pairs. Each pair included a high-constraint sentence and a low-constraint sentence (based on a cloze probability questionnaire, as detailed below). See Table 1 for an example set and Appendix B for all materials. The sentences in each pair differed at the beginning but were identical from the critical words onwards. The critical words were the verb after which the sentence constraint differed between the conditions, and the NP that followed it. In order to avoid contaminating the ERPs measured on the NP by effects stemming at the verb, these words were separated by the word '*et*', the Hebrew accusative case marker. This case marker does not bear any agreement features, and it was therefore identical in all sentences and could not provide any indication of the upcoming noun. The number of words before the verb did not differ between the conditions ($p = .890$), nor did the length or frequency of the word prior to the verb ($p = .163$ and $p = .383$, respectively). Materials were divided into two lists according to a Latin square such that each participant saw only one sentence from each pair (i.e. 26 items in each condition, and 52 sentences in total). Presentation order was randomized for each participant.

Table 1: Example set

Constraint	Sentence
High	*Biglal še-ofir lo makir et ha-sifria, ha-safranit azra* since that-ofir not know ACC the-library, the-librarian helped *lo* ***limco*** *et* ***ha-sefer*** *še-hu haia carix* him **to-find** ACC **the-book** that-he COP needed 'Since Ofir isn't familiar with the library, the librarian helped him **find the book** he needed'
Low	*ofir xipes ve-xipes bemešex šaot, aval lo ecliax* ofir searched and-searched for hours, but not succeeded ***limco*** *et* ***ha-sefer*** *še-hu haia carix* **to-find** ACC **the-book** that-he COP needed 'Ofir had searched for hours, but he couldn't **find the book** he needed'
Yes/No question: Was Ofir looking for a book? (Yes)	

The critical words (the verb and the NP) are marked in bold. ACC = accusative case, COP = copula.

Cloze probability questionnaires: Two cloze probability questionnaires were conducted. Both included sentence fragments, and participants were instructed to complete each sentence with the first completion that comes to mind. In the first questionnaire, the sentences were presented truncated after the verb (i.e. the presented sentence frame included the verb). This questionnaire included 66 sentence pairs, divided into two lists such that each participant saw only one sentence from each pair. The order of presentation was randomized for each participant. A hundred and two participants completed this questionnaire (average age 25.4, 29 male). Based on this questionnaire, the 52 experimental sets were chosen. The average constraint in the high-constraint condition was 72.4% (range: 50% - 100%), meaning that on average, 72.4% of completions provided by participants in the questionnaire were 'et' (the accusative case marker) followed by the most commonly provided noun. The average constraint in the low-constraint condition was 18.8% (range: 0-50%, again, percentage reflecting 'et' + the most commonly provided noun). When counting completions of the same lexical item whether it was preceded by the accusative case marker or not, the average constraint in the high-constraint condition was 80.5% and the average constraint in the low-constraint condition was 27.3%. Overall, the accusative case mark itself (with any noun) was predictable in both conditions (high – 87.0%, low – 74.8%).

In the second cloze probability questionnaire, the sentences were truncated before the verb. Since the verb is a critical word in the experiment (i.e. its ERP is of interest), this questionnaire was aimed to make sure that the verb's cloze probability did not differ between conditions, which would lead to an N400 effect. The questionnaire included the 52 experimental sets, divided into two lists such that each participant saw only one sentence from each pair. The order of presentation was randomized for each participant. This questionnaire was also completed by a hundred and two participants, different from the ones who completed the first questionnaire (average age 24.3, 25 male). The average cloze probability of the verbs was 0.7% and 1.1% in the high- and low-constraint conditions respectively, with no significant difference between conditions ($p = .595$).

2.3) Procedure

Stimuli were presented using the E-prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Sentences were presented word-by-word in the middle of the screen for 250ms, with a 350ms ISI. A comprehension question appeared following 50% of the trials (randomly distributed). Each trial was preceded by a 1000ms fixation point. After each trial a string of number signs (####) appeared on the screen and the participant pressed a button to start the next trial. Participants were encouraged to take as many breaks as needed. Prior to the experiment, participants completed a practice block of five trials.

Reading span test: To asses WM capacity, each participant also completed a reading span test.[4] The test was performed following the main experiment or, for a few participants, in a separate session prior to their participation in the main experiment. The test's procedure is based on Daneman and Carpenter (1980), with minor differences. Participants read aloud a series of Hebrew sentences, after which they had to recall the last word of each sentence. The number of sentences in the series increased from two to six. Participants had three series in each level, and the last level at which a participant correctly recalled all words in at least two series was defined as this participant's reading span (i.e., when the participant failed to recall a word in two series in the same level the test was terminated and the participant's reading span was defined at the prior level). Two practice series (at the two-sentence level) were performed prior to the test, in which participants could make mistakes and ask questions.

2.4) EEG recording and pre-processing

The electroencephalogram (EEG) was recorded using a BrainVision actiCap system with 32 Ag/AgCl scalp electrodes attached according to the 10-20 system. Two electrodes were used to monitor EOG, located at the outer canthi and the infraorbital ridge of the right and left eyes respectively. Electrode impedances were kept below 5 kΩ for all scalp electrodes and below 15Ω for the EOG electrodes. During recording, the EEG was referenced to Fp2 for most participants (the online reference electrode for five participants was Fp1, for technical reasons). The EEG was then re-referenced offline (for all participants) to the average of the left and right mastoid electrodes. Data were collected at a 250 Hz sampling rate and

[4] Reading span is a limited measure of WM capacity, as it can also be affected by differences in language proficiency. Participants in the current study were likely rather homogenous in their language proficiency since they were all mono-lingual Hebrew-speaking university students, which would allow reading span scores to reflect WM capacity relatively well. However, future studies can disentangle WM capacity from language proficiency, possibly by using a factor analytic approach (e.g. Kim, Oines, and Miyake, 2017).

low-pass filtered at 70 Hz. Data were then bandpass-filtered between 0.1 and 30 Hz, and segmented into 1200 ms epochs, including -200 to 1000 ms relative to the onset of the critical word. The 200 ms prior to the onset of the critical word were used for baseline correction. Trials contaminated by blinks, eye movements, excessive muscle activity or amplifier blocking were rejected off-line before averaging and excluded from further analysis (this affected 5.36% of the trials).

2.5) EEG data analysis

Based on the typical time windows of the N400 and the P600, mean amplitudes over 300-500ms and 500-800ms (respectively) were analyzed. Electrodes were grouped based on their anteriority and laterality (Anterior - Left: F7, F3, Fp1, FC5, FC7, T1, C3; Middle: Fz, Cz; Right: F8, F4, Fp2, FC2, FC6, T7, C4; Posterior - Left: P7, P3, O1, CP5, CP1; Middle: Pz, Oz; Right: P8, P4, O2, CP2, CP6) in order to reduce the number of comparisons and the familywise error rate (see Luck, 2014) while still allowing to assess the topography of the effects. Standardized reading span scores were entered to the analyses as a continuous covariate. This resulted in repeated-measures ANCOVAs with the factors Anteriority (Anterior, Posterior), Laterality (Left, Middle, Right), and Constraint (High, Low), and the covariate Span. These analyses were conducted on time windows relative to both the verb onset and the noun onset, and were followed by separate analyses for anterior and posterior sites (with the factors Laterality and Constraint, and the covariate Span). The Huyhn-Feldt adjustment for nonsphericity of variance was applied when the sphericity assumption was violated. In these cases, the corrected p-value is reported with the original degrees of freedom.

3) Results

3.1) Accuracy

Accuracy for the comprehension questions was significantly above chance for all participants. Mean accuracy rate was 93.37% (SD: 4.83). Accuracy data were subjected to a repeated-measures ANCOVA with the factor Constraint (High, Low) and the covariate Span. No significant main effects or interactions were found.

3.2) EEG

Grand averaged ERPs and scalp distributions of the components are displayed in Figures 2-4. The results of the different ANCOVAs are provided in Table 2.

<u>Verb</u>

N400: Mean amplitudes over the 300-500ms time window (relative to the verb onset) were entered into a repeated-measures ANCOVA with the factors Anteriority (Anterior, Posterior), Laterality (Left, Middle, Right), and Constraint (High, Low), and the covariate Span, followed by separate ANCOVAs for anterior and posterior sites. No main effects or interactions were found.

P600: Mean amplitudes over the 500-800ms time window (relative to the verb onset) were entered into a repeated-measures ANCOVA with the factors Anteriority (Anterior, Posterior), Laterality (Left, Middle, Right), and Constraint (High, Low), and the covariate Span. A significant interaction was found between Anteriority and Constraint ($F(1,34) = 4.312$, $p = .045$), as well as between Anteriority, Constraint and Span ($F(1,34) = 6.267$, $p = .017$). To further explore these interactions, separate ANCOVAs were performed on posterior and anterior sites. In the ANCOVA conducted over posterior electrodes, a significant effect of Constraint was found ($F(1,34) = 11.935$, $p = .001$) such that P600 amplitude was higher in the high-constraint condition. Additionally, a significant interaction was found between Constraint and Span ($F(1,34) = 5.698$, $p = .023$), such that the P600 effect was grater for participants with a higher span. The relation between reading span and the P600 effect at the verb is plotted in Figure 5. These effects were not found in the ANCOVA conducted over anterior electrodes.

<u>Noun</u>

N400: Mean amplitudes over the 300-500ms time window (relative to the noun onset) were entered into a repeated-measures ANCOVA with the factors Anteriority (Anterior, Posterior), Laterality (Left, Middle, Right), and Constraint (High, Low), and the covariate Span. A significant interaction was found between Anteriority and Constraint ($F(1,34) = 15.024$, $p < .001$) but not between Anteriority, Constraint and Span. To follow-up on the two-way interaction, separate ANCOVAs were run on posterior and anterior sites. In the ANCOVA conducted over posterior electrodes, a significant effect of Constraint was found ($F(1,34) = 19.057$, $p < .001$), such that N400 amplitude was higher in the low-constraint condition. There was no interaction between Constraint and Span. Neither of these effects was found in the ANCOVA conducted over anterior electrodes.

P600: Mean amplitudes over the 500-800ms time window (relative to the noun onset) were entered into a repeated-measures ANCOVA with the factors Anteriority (Anterior, Posterior), Laterality (Left, Middle, Right), and Constraint (High, Low), and the covariate Span. There was no significant interaction between Anteriority and Constraint, but we did observe a trend towards a three-way interaction between Anteriority, Constraint and Span ($F(1,34) = 2.883$, $p = .099$). Although this interaction did not reach significance, we further explored this pattern by conducting separate ANCOVAs over posterior and anterior sites, since our a-priory hypothesis was about posterior effects (the P600 component). In the ANCOVA conducted over posterior electrodes, no significant effect of Constraint was found, but there was a significant interaction between Constraint and Span ($F(1,34) = 4.422$, $p = .043$), such that for participants with higher reading span, there was a greater decrease in P600 amplitude in the high- (relative to low-) constraint sentences. This effect was not found in the ANCOVA conducted over anterior electrodes. There was also a significant three-way interaction between Laterality, Constraint and Span ($F(2,68) = 5.57$, $p = .011$), suggesting that the observed P600 was left-lateralized.

To summarize the results, at the verb, a significant effect of Constraint as well as an interaction between Constraint and Span were found in posterior electrodes in the 500-800ms time window, meaning that high-constraint sentences elicited an increased P600 amplitude (relative to low-constraint ones), and that this increase was greater for participants with higher WM capacity.

At the noun, a significant effect of Constraint was found in posterior electrodes in the 300-500ms time window, with high-constraint sentences eliciting a decreased N400 amplitude (relative to low-constraint ones). This N400 effect was not affected by WM capacity. Additionally, an interaction between Constraint and Span was found in posterior electrodes in the 500-800ms time window on the noun, meaning that for participants with higher reading span, there was a greater decrease in P600 amplitude in the high- (relative to low-) constraint sentences.

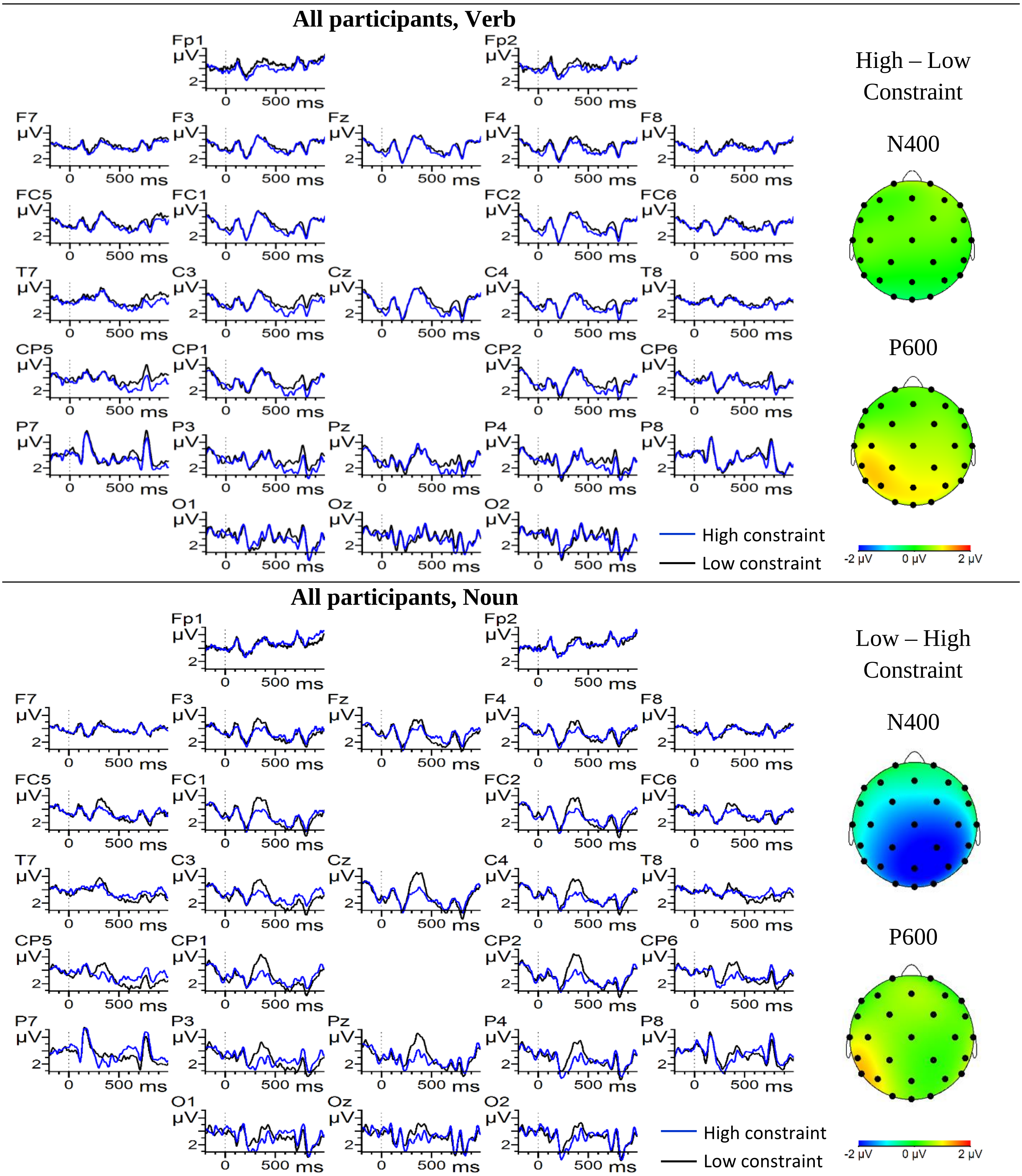

Figure 2: Grand averaged ERPs and scalp distributions for all participants. Grand averaged ERPs and scalp distributions of the N400 (300-500ms) and P600 (500-800ms) of the verb and noun for all participants.

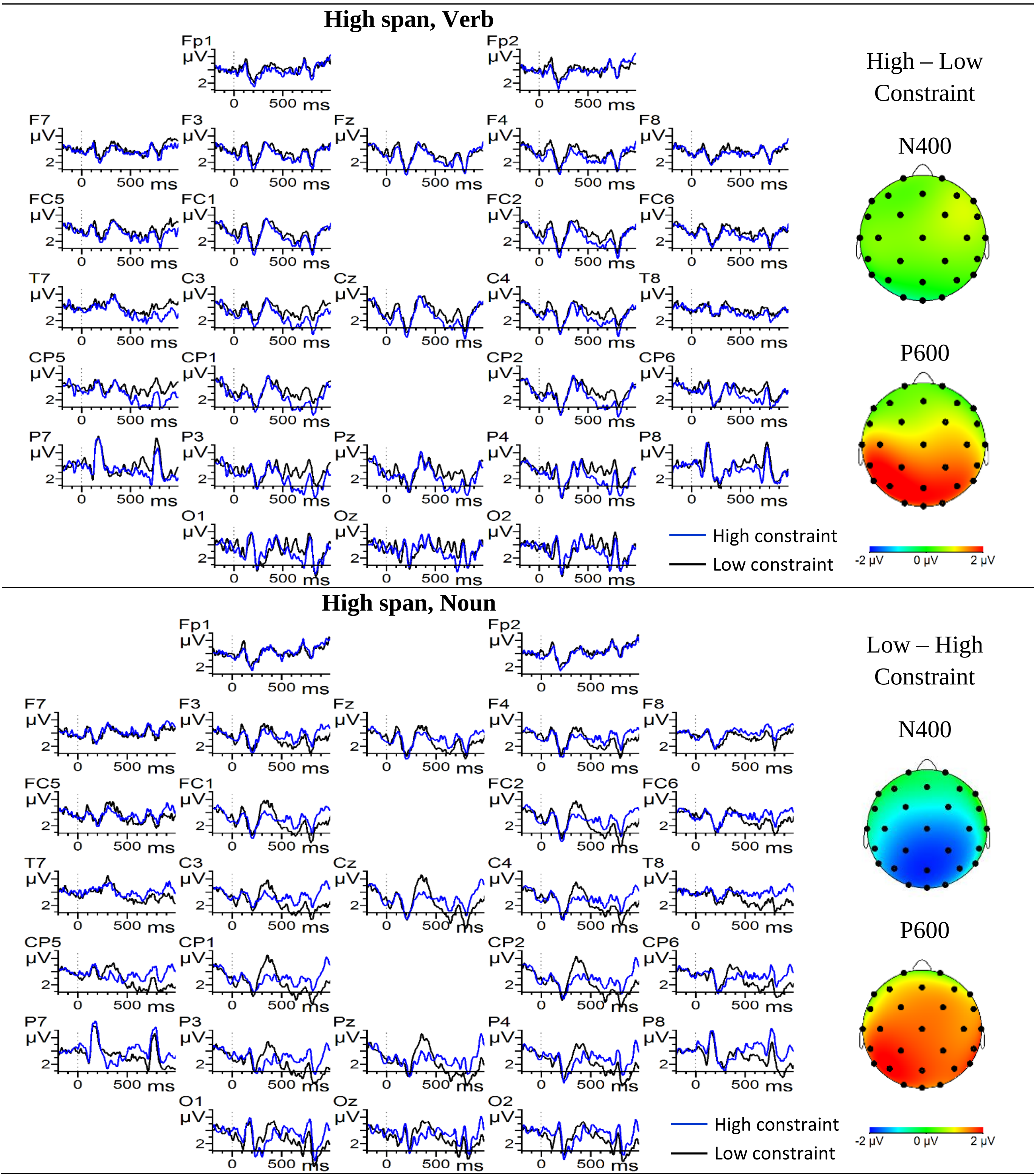

Figure 3: Grand averaged ERPs and scalp distributions for the high-span group. Grand averaged ERPs and scalp distributions of the N400 (300-500ms) and P600 (500-800ms) of the verb and noun for high span participants, with a reading span of 4 or higher (M = 4.33).

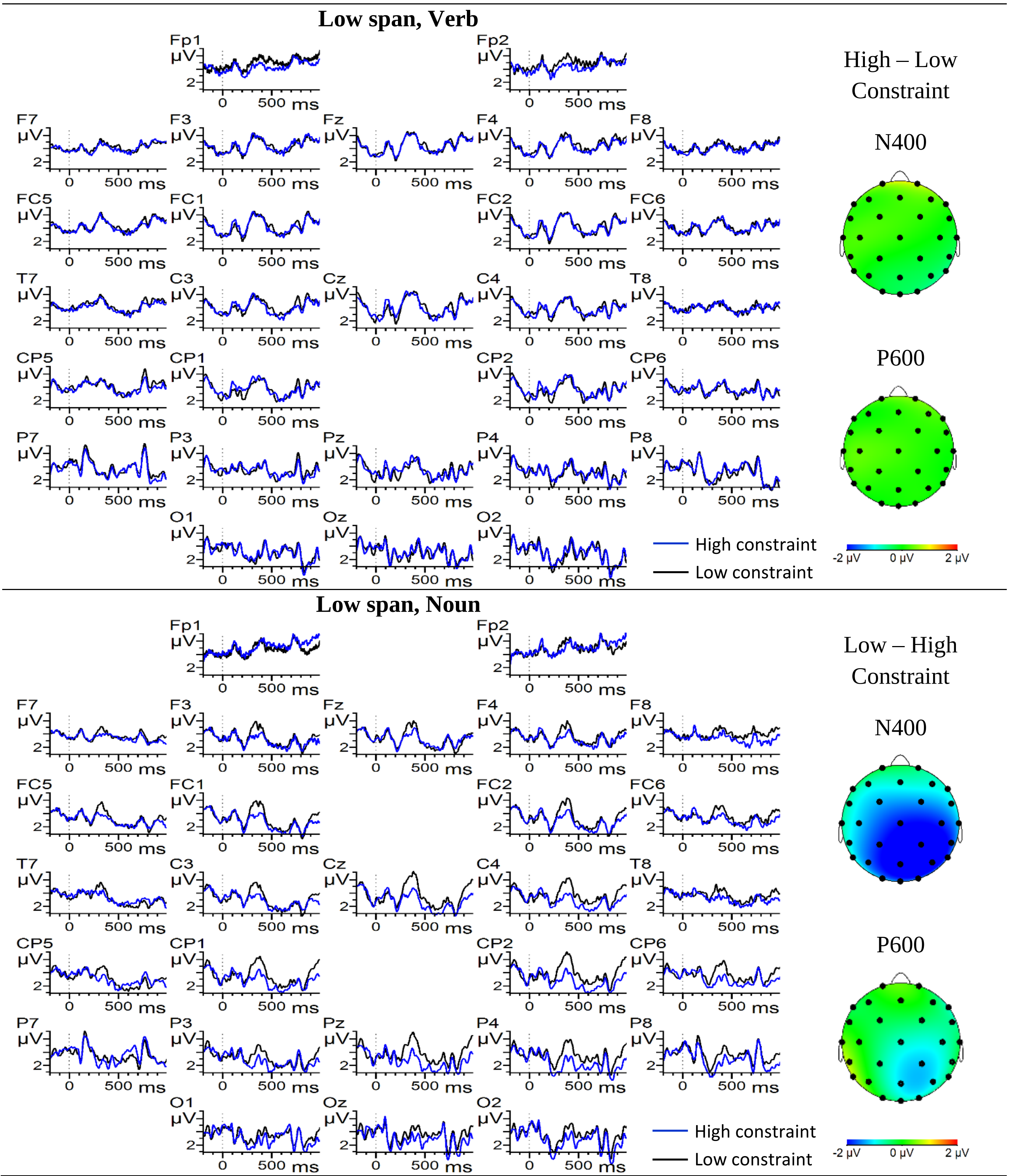

Figure 4: Grand averaged ERPs and scalp distributions for the low-span group. Grand averaged ERPs and scalp distributions of the N400 (300-500ms) and P600 (500-800ms) of the verb and noun for low span participants, with a reading span of 3 or lower (M = 2.61).

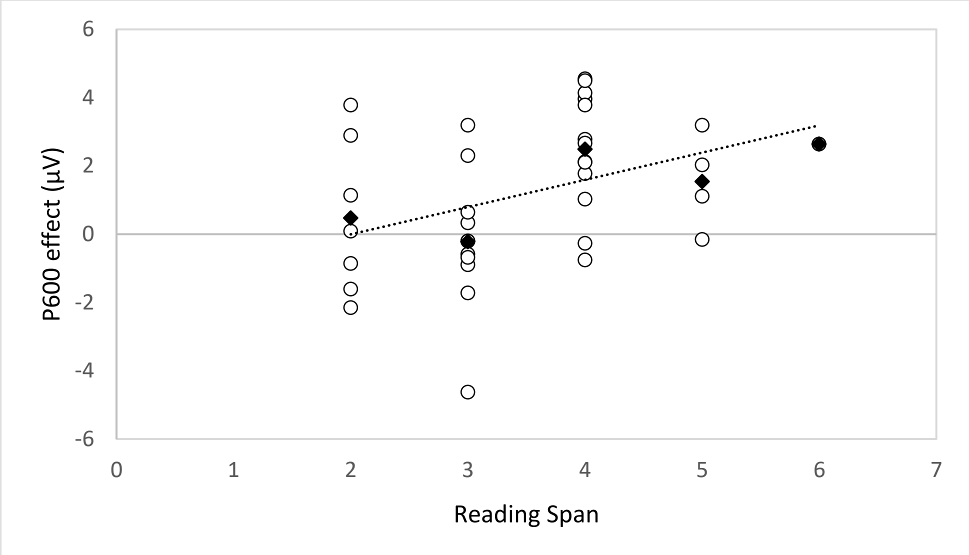

Figure 5: Size of P600 effect at the verb as a function of reading span score. The size of the P600 effect is calculated as the difference in mean amplitude between the high and low constraint conditions over the 600-800ms time window (relative to verb onset), in posterior electrodes. ○ – single participant, ♦ - average across reading span score

Table 2: Results of ANCOVAs on EEG data

		Verb				Noun			
		N400		P600		N400		P600	
	df	F	p	F	p	F	p	F	p
Constraint	1,34	0.03	.866	8.45	.006	13.41	.001	1.81	.187
Constraint*Span	1,34	0.03	.827	2.12	.154	0.10	.753	3.66	.064
Anteriority*Constraint	1,34	1.51	.227	4.31	**.045**	15.02	< **.001**	0.01	.921
Anteriority*Constraint*Span	1,34	0.07	.796	6.27	**.017**	2.08	.158	2.88	**.099**
Laterality*Constraint	2,68	0.62	.509	0.72	.491	4.82	.014	2.24	.119
Laterality*Constraint*Span	2,68	1.84	.175	0.48	.620	3.56	.038	6.76	.003
Anteriority*Laterality*Constraint	2,68	0.36	.681	0.36	.662	0.72	.486	4.75	.015
Anteriority*Laterality*Constraint*Span	2,68	0.20	.796	0.04	.942	0.69	.499	0.22	.772

		Verb							
		N400				P600			
		Anterior		Posterior		Anterior		Posterior	
	df	F	p	F	p	F	p	F	p
Constraint	1,34	0.57	.455	0.25	.623	2.81	.103	11.94	< **.001**
Constraint*Span	1,34	0.71	.792	0.00	.994	0.03	.876	5.70	**.023**
Laterality*Constraint	2,68	0.23	.796	1.08	.336	0.56	.576	0.73	.486
Laterality*Constraint*Span	2,68	0.73	.488	2.58	.093	0.31	.731	0.41	.668

	Noun								
		N400				P600			
		Anterior		Posterior		Anterior		Posterior	
	df	F	p	F	p	F	p	F	p
Constraint	1,34	2.56	.119	19.06	**< .001**	2.16	.151	1.18	.282
Constraint*Span	1,34	0.30	.590	0.72	.404	1.65	.208	4.42	**.043**
Laterality*Constraint	2,68	4.42	.043	3.10	.064	0.58	.563	6.21	.007
Laterality*Constraint*Span	2,68	1.36	.264	5.09	.014	4.24	.018	5.57	.011

Factors: Anteriority (Anterior, Posterior), Laterality (Left, Middle, Right), Constraint (High, Low). Covariate: Span. N400 time window: 300-500 ms. P600 time window: 500-800 ms. The table includes all results involving the manipulated factor, Constraint (i.e. the main effect and any interaction that includes Constraint).

4) Discussion

In this study, we looked for evidence in support of the pre-updating process, namely, the integration of predicted material into the sentence representation in working memory prior to its occurrence in the input. Our results showed increased P600 amplitude at the verb in the strongly constraining sentences, and this effect was greater for participants with higher working memory span. We propose that this effect reflects integration processes prior to the onset of a highly predictable word, thus providing evidence in support of pre-updating. The benefits of pre-updating were observed on the predictable NP, where participants with higher reading span showed a greater decrease in P600 amplitude in the strongly constraining sentences (relative to the weakly constraining sentences), suggesting that the integration of an already pre-updated word was easier. These effects had the posterior distribution typical for the P600 component. It can be mentioned that a late frontal positivity would not be expected here, as this positivity is elicited by unexpected words in highly constraining sentences (e.g. Brothers, Swaab, Traxler, 2015; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007), and is therefore thought to be related to coping with disconfirmed predictions (see further discussion of this anterior positivity below).

Additionally, an N400 effect was found at the NP in the low-constraint sentences relative to the high-constraint sentences. This result replicates the very robust finding that the amplitude of the N400 is greater for words with low cloze-probability relative to words with high cloze-probability (e.g. Delong, Urbach, & Kutas, 2005; Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984; Wlotko & Federmeier, 2012). This N400 effect did not differ between the high- and low-span participants. The fact that the P600 effects were affected by WM capacity while the N400 effect was not, taken together with previous literature regarding the different processes reflected by these two components, corroborates the suggestion that although both pre-activation and pre-updating are expected to occur more often in the high-constraint condition, we can see two distinct neural responses that behave differently, reflecting the two distinct processes. The P600 effects being affected by WM capacity is in line with the hypothesis that pre-updating involves WM representations, while pre-activation occurs in long-term memory. These results support an architecture whereby what drives the differences due to WM capacity is not a difference in activation (as this would likely manifest as a difference in N400 amplitude) but rather a difference in the retrieval threshold which determines the activation level sufficient for pre-updating. As mentioned in the Introduction, we see pre-updating as a process that is initiated only if the activation of the predicted content exceeds a certain threshold. The difference seen in the current study between participants with different reading spans can be interpreted as an indication that this threshold is generally lower for participants with a higher WM capacity, since pre-updating is less costly for them. Such an architecture would result in similar activations for all participants, but different likelihood to pre-update, in line with the current results.

4.1) Previous studies on the processing of predictive words

Several recent studies have investigated processing at the stage prior to the occurrence of a predictable word. Li, Zhang, Xia, and Swaab (2017) compared high- and low-constraint sentences in Mandarin Chinese, looking at the verb preceding a predictable or an anomalous noun. Their results did not show the P600 effect observed in our experiment, but rather a sustained anterior negativity (SAN) elicited at the verb, as well as reduced beta power (19-25 Hz). There are several crucial differences between the current experiment and the experiment of Li et al. that may have led to these different results. The first and probably most influential difference is that the critical region in the Li et al. experiment consisted of a verb, followed by a classifier or an adjective, followed by a noun. While the noun was predictable immediately after the verb, based on a sentence completion questionnaire, the classifier/adjective was not predictable. This means that in every trial the participant encountered an unpredictable word immediately following the verb. This was not the case in our study, as the verb in our sentences was always followed by the Hebrew accusative case marker 'et', a function word that was predicted based on the cloze-probability questionnaire (and possibly even more predictable in the experimental context as it consistently appeared in all sentences), which was followed by the noun. This allowed participants in the current study to precisely predict the direct object of the verb (i.e. 'et' + noun, forming the NP that immediately follows the verb), enabling its immediate integration. Another difference between the Li et al. experiment and ours is that since the authors of the former experiment aimed to also test the consequences of prediction failure, the critical noun in their experiment was anomalous in half of the trials, leading to a disconfirmed prediction. It is conceivable that the proportion of disconfirmed predictions in the experimental context affects participants' tendency to commit to a prediction by pre-updating it, namely that if participants repeatedly have to endure the cost of prediction failure they would become more cautious and avoid pre-updating. Namely, the retrieval threshold that initiates pre-updating may not only vary between individuals, but also be adjusted via top-down control to adapt to different situations. In situations with a strong incentive to predict, the threshold would be lowered, leading to more frequent occurrence of pre-updating. In situations such as Li et al.'s experiment, the threshold would be raised in order to prevent commitment to prediction which would likely be disconfirmed. This means that although predicted candidates would still be pre-activated, pre-updating would be less likely to occur. These differences may explain why no P600 effect was found at the Li et al. study. Moreover, if participants in that study noticed that they consistently encounter an unpredictable word immediately after the verb, but that the predictable word may still appear afterwards, then the results may reflect holding a (not-yet-integrated) prediction that is not expected to be realized immediately. This aligns well with the suggestion that the SAN, sometimes observed between the filler and gap sites in long-distance dependencies, is an index of working memory load (e.g. King & Kutas, 1995; Phillips, Kazanina, & Abada, 2005). Additionally, in Li et al. the influence of WM capacity was not tested. It is possible that due to random sampling, the participants in that study were on average with lower WM capacity then our participants, and as the effects seen in our experiment were driven by the high-span participants, this could further contribute to the difference in the results between the two studies.

Rommers and colleagues (2017) have also looked at the stage prior to the occurrence of a predictable word. They performed a reanalysis of data from a study comparing high- and low-constraint sentences, ending with the most predictable word or with an unexpected word (Federmeier et al., 2007). While in the original study, ERPs of the final word were analyzed, the new analysis focused on the time-frequency domain, looking at the final word as well as a time-window prior to its onset. Similar to Li et al., Rommers et al. also found an effect of constraint in the time-frequency domain prior to the predicted (or unexpected) word. However, the results of Rommers et al. show a decrease in alpha power (8-12 Hz), rather than in beta power, in the high-constraint sentences. These inconsistent results highlight the fact that in investigating predictive processes we must carefully consider factors that are likely to affect the nature of predictions, such as predictive validity within the experimental context (i.e. the proportion of trials in which predictions are disconfirmed), the immediacy of the predicted content, etc.

Fruchter and colleagues (2015) conducted a MEG experiment investigating processing of the adjective in adjective–noun phrases, when the adjective was predictive of the upcoming noun to different degrees (e.g. 'stainless' is highly predictive of the noun 'steel', whereas 'important' is not predictive of any particular noun). They found greater activity in the left middle temporal gyrus for highly predictive adjectives, and a significant interaction in the same area between adjective predictivity and the frequency of the expected noun, such that higher noun frequency led to decreased activity when the adjective was very predictive. We believe that this interaction with the noun's frequency indicates that the effect found in that experiment is more likely to reflect pre-activation then pre-updating, as pre-activation is affected by lexical properties of the word being activated, while pre-updating occurs after the word has already been retrieved (i.e. fully activated) and is therefore not expected to be affected by the lexical properties of the word.

To test whether lexical properties of the predicted words also underlie the P600 effect seen on the verb in our experiment, we performed a post-hoc by-items correlation analysis between the log frequency of the predicted word (taken from the corpus of Linzen, 2009) and the average amplitude of the P600 effect on the verb (defined as the difference between the high- and low-constraint conditions, in the 500-800ms time-window, in the 9 centro-parietal and parietal electrodes). No significant correlation was found (Pearson's $r = .076$, $n = 52$, $p = .594$). We acknowledge that this is a null result and therefore no strong conclusions can be drawn from it. However, this result is compatible with our conjecture that pre-activation, but not pre-updating, would be affected by lexical properties of the predicted word.

4.2) Individual WM differences and the P600

In the current study, the P600 effects associated with pre-updating were greater for participants with higher reading span, indicating that participants with higher WM capacity are more likely to engage in this process. A couple of previous studies found that the P600 component was affected by individual differences in WM capacity. Nakano and colleagues (2010) manipulated the first NP in simple SVO sentences such that the verb following it would either be plausible, implausible due to world knowledge, or in violation of thematic requirements (i.e. an inanimate subject when the verb requires an animate one). Their results showed that violation of world knowledge led to a similar N400 effect regardless of WM capacity. However, thematic violations led to distinctly different responses in the low- and high-span groups. An N400 effect (see discussion in the next subsection) was only observed in the low-span group, while a P600 effect was observed in the high-span group. The P600 effect in the high-span group may indicate that participants in this group made predictions regarding the upcoming verb based on the subject's properties (i.e. its inanimacy), and therefore had greater difficulty integrating a verb requiring an animate agent.

Vos and Friederici (2003) showed that in locally ambiguous sentences, disambiguation towards the less expected syntactic structure elicits a P600 effect at the disambiguating word only for participants with high WM capacity. This effect may indicate that high-span participants had to perform a syntactic reanalysis when the more predictable structure was incorrect since they had made structural predictions prior to the disambiguating word. One possible commonality between the findings in the two studies is that differences in the P600 component between high- and low-span participants tend to arise when the cause of the P600 effect is not syntactic difficulty per se (i.e. a complex or ungrammatical structure) but rather difficulty that stems from an incorrect prediction (albeit this would not necessarily be the case for all effects of WM capacity on the P600 component, see Kim, Oines, & Miyake, 2018). The cause of such differences may therefore be increased engagement in predictive processing by individuals with high WM abilities, stemming from the fact that these individuals have the available resources needed for the prediction itself, as well as sufficient resources to endure the costs incurred in the case of a contradictory input.

4.3) Individual WM differences and the N400

Several studies have tested the influence of individual differences in WM capacity on the N400 component. For example, Van Petten and colleagues (1997) compared the use of lexical context and sentence-level context by participants with low-, mid- and high-WM capacity. Pairs of associated or unassociated words were embedded in congruent and incongruent sentences. Similar N400 effects for participants with low-, mid- and high-WM capacity were observed when comparing associated and unassociated word pairs, but the contrast between unassociated word pairs in congruent and incongruent sentences revealed a significant N400 effect only in participants with mid- and high-WM capacity. These results were interpreted as indicating that WM capacity affects the degree to which purely sentence-level context is used for prediction, but not the degree to which lexical context is. A more qualitative difference in ERP responses of participants with high- and low-WM capacity was shown by Nakano and colleagues (2010). As explained in the previous subsection, in this experiment, thematic violations elicited an N400 effect only in the low-span group, while a P600 effect was observed in the high-span group. World knowledge violations led to similar N400 effects in both groups.

The generalization that seems to emerge from these two studies is that the amplitude of the N400 is not affected across the board by individual differences in WM capacity. Rather, specific types of information (i.e. sentence context, thematic requirements) can be differentially used for pre-activation by readers with higher or lower working memory span, leading to differences in the N400 effect. In contrast, effects that stem from lexical context are not affected by WM capacity. We propose that this is consistent with the fact that in the current experiment the N400 effect did not differ between the two participant groups. This is so since in our materials, the decrease in the amplitude of the N400 in the high-constraint sentences could have originated solely from their lexical content, facilitating retrieval of the predicted word via spreading activation. To test this proposal, we conducted a questionnaire in which participants were given lists of words, with each list containing only the content words of one experimental sentence, randomly ordered (i.e. not in the same order as in the original sentence).

Participants were asked to read each list and provide the first association that comes to mind. The results of this questionnaire showed that the percentage of participants providing the experimental critical NP as an association for the word list was significantly greater in the high-constraint sentences then in the low-constraint ones (the average percentage was 3.4% and 37.6% for the low- and high-constraint sentences respectively, $t(52) = 10.52$, $p < .001$). We do not claim that these results indicate that in our experiment only lexical context was used for prediction and sentence-level context was ignored. However, the results do indicate that an N400 effect would be expected in low-span participants even if these participants relied mostly (or only) on lexical context for prediction.

4.4) Pre-updating and prediction failure

As mentioned in the Introduction, pre-updating entails some form of commitment to a prediction. However, the integration carried out during pre-updating may not be identical to the integration occurring upon actually encountering a word in the input, which includes building syntactic structure, assigning thematic roles, integrating the semantic content of the word, etc. If the integration of a predicted word was identical to that of an actual word, disconfirming a pre-updated prediction would lead to a bona fide reanalysis. On the other hand, if the integration of a predicted word is tentative, or not as complete as the integration of an actually encountered word, then disconfirmation of such prediction, though still costly, would not entail a typical reanalysis.

Previous ERP studies have shown that a congruent-unexpected word that appears instead of a highly predictable word elicits a frontal positivity termed the "frontal post-N400 positivity" (e.g. Delong, Urbach, Groppe, & Kutas 2011; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Ness & Meltzer-Asscher, in press). The frontal distribution of this component is distinct from the distribution of the P600 that is commonly elicited when a syntactic reanalysis is performed (e.g. Osterhout, Holcomb, & Swinney, 1994; Hagoort, Brown & Osterhout, 1999), and it was suggested to reflect discourse revision (i.e. increased difficulty in updating a context after it was already wrongly updated, Brothers, Swaab, & Traxler, 2015) or inhibition of the falsely predicted word (Ness & Meltzer-Asscher, in press). The fact that a frontal positivity rather than a P600 is elicited in such cases may indicate that a pre-updated prediction has a different status then an actually encountered word, displaying tentative integration which can more easily be undone.

4.5) An alternative approach – Surprisal and Entropy reduction

Throughout the paper, the hypotheses and the interpretation of the results were framed under the assumption of serial parsing; however, our results can also be conceptualized in the framework of parallel and probabilistic parsing, which relate processing difficulty throughout a sentence to complexity matrices reflecting the amount of information conveyed in each word. One common matrix of this kind is Surprisal, reflecting the probability of a word given the preceding words in the sentence (Hale, 2001, Levy, 2008). Many studies have shown that greater surprisal is associated with increased processing difficulty (e.g. Boston, Hale, Kliegl, Patil, & Vasishth, 2008; Demberg & Keller, 2008; Frank, Otten, Galli, & Vigliocco, 2015; Smith & Levy, 2013). Another relevant matrix is Entropy reduction. Entropy reflects the degree of uncertainty regarding what is being conveyed, and it is therefore high when many possible outcomes have similar probabilities, and lower when there is a very probable outcome. As each word is encountered, it affects expectations and the probability distribution changes. The degree to which a given word reduces uncertainty (i.e. reduces entropy) represents the amount of information gained

(Hale, 2003). Processing more information incurs greater processing difficulty, and indeed, studies have shown that greater entropy reduction leads to increased reading times, independently of the contribution of surprisal (Frank, 2013; Linzen & Jaeger, 2016; Wu, Bachrach, Cardenas, & Schuler, 2010. See Hale, 2016 for a review).

We can now consider our materials in light of these findings. At the verb, cloze probability did not differ between conditions, and therefore surprisal was similar. However, while entropy was always relatively high prior to the verb, certainty regarding the noun was much greater in the high-constraint condition than in the low-constraint one. This means that entropy reduction at the verb was greater in the high-constraint condition, which should lead to greater processing difficulty. At the noun, cloze probability was greater in the high-constraint condition than in the low-constraint one. Therefore, surprisal was lower in the high-constraint condition, which would lead to decreased processing difficulty (relative to the low-constraint condition). Regarding entropy reduction at the noun, while, as explained above, entropy prior to the noun was lower in the high-constraint condition than in the low-constraint one, entropy following the noun was likely high in both conditions (as in both conditions there was no predicted continuation for the sentence after the noun). This means that at the noun there was no entropy reduction in any condition (entropy either increased or did not change, both considered the same in terms of processing difficulty, see Hale, 2016; Lowder et al., 2018). To recap, in the high- (relative to the low-) constraint condition, increased processing difficulty would be predicted at the verb due to entropy reduction (surprisal being similar in both conditions), and decreased processing difficulty would be predicted at the noun due to lower surprisal (entropy reduction likely being similar in the two conditions). These predictions are in line with our results, as the increased P600 on the verb in the high-constraint condition may be taken to index greater entropy reduction, and the decreased N400 and/or P600 on the noun in the high-constraint condition may be taken to reflect lower surprisal.

Additional work is needed in order to provide an explanation within this framework as to why the observed P600 effects depended on WM capacity. Possibly, this can be done by assuming that individuals with higher WM capacity have a larger “beam size”, namely they consider more possible continuations, which would vary the Entropy experienced by different individuals. Additionally, making predictions regarding the specific ERP components which reflect surprisal and entropy reduction is not trivial and further research is needed to account, within this framework, for why we see a P600 effect at the verb, while at the noun we see both N400 and P600 effects (e.g. see Cho et al., 2018 for a mechanistic account for surprisal. Such accounts can provide a means to make predictions regarding timing and neural activity).

4.6) Prediction mechanisms in sentence processing

Figure 6 summarizes the proposed prediction mechanisms during sentence processing, and how they are incorporated within the processing stages of a word appearing in sentence context. At every stage during sentence comprehension, multiple representations in long-term memory are pre-activated. The activation level of a word is affected by the previous context, as well as by the lexical properties of that word. Top-down control may also limit or enhance activations, mediating the influence of factors such as predictive validity, task demands, etc. Once the activation level of a certain word had passed a retrieval threshold, it is regarded as retrieved, i.e. it can be integrated into the representation being built in WM. This threshold may differ between individuals and it may also be adjusted by top-down control, adapting to how beneficial commitment to predictions is in a given situation (e.g. due to task demands, noisy input, predictive validity).

This means that top-down control can affect prediction in two ways: first, by limiting or enhancing activation levels, which enables differential influence on different representations. It is conceivable that this kind of adjustment would take place when a certain information type is more or less reliable in a given situation. For example, if when reading a fantasy book animacy requirements are often violated, then animacy may be less heavily relied upon for prediction and therefore words that violate an animacy requirement may be predicted and activated. Second, top-down control can affect prediction by raising or lowering the retrieval threshold, which enables a more general influence on the likelihood of any predicted content to reach retrieval prior to its realization in the input. This kind of adaptation will take place for example when predictive validity is low, meaning that predictions are often violated and therefore forming strong predictions is not beneficial.

If no word had passed the retrieval threshold prior to the realization of the next word in the input, then bottom-up activation causes retrieval of the word, and it is then integrated into the representation in WM. If a certain word did pass the threshold prior to the realization of the next word in the input, this word would then be pre-updated, meaning that a tentative integration would be initiated, until the pre-updated word can be matched against the input. At this point, if the input matches the pre-updated word, integration is finalized. This stage would be less demanding than the integration of a word that had not been pre-updated. If the input does not match the pre-updated word, then the falsely predicted word is inhibited in order to enable integration of the actual input. This mismatch between prediction and input may also be a valuable trigger for learning mechanisms, improving future predictions.

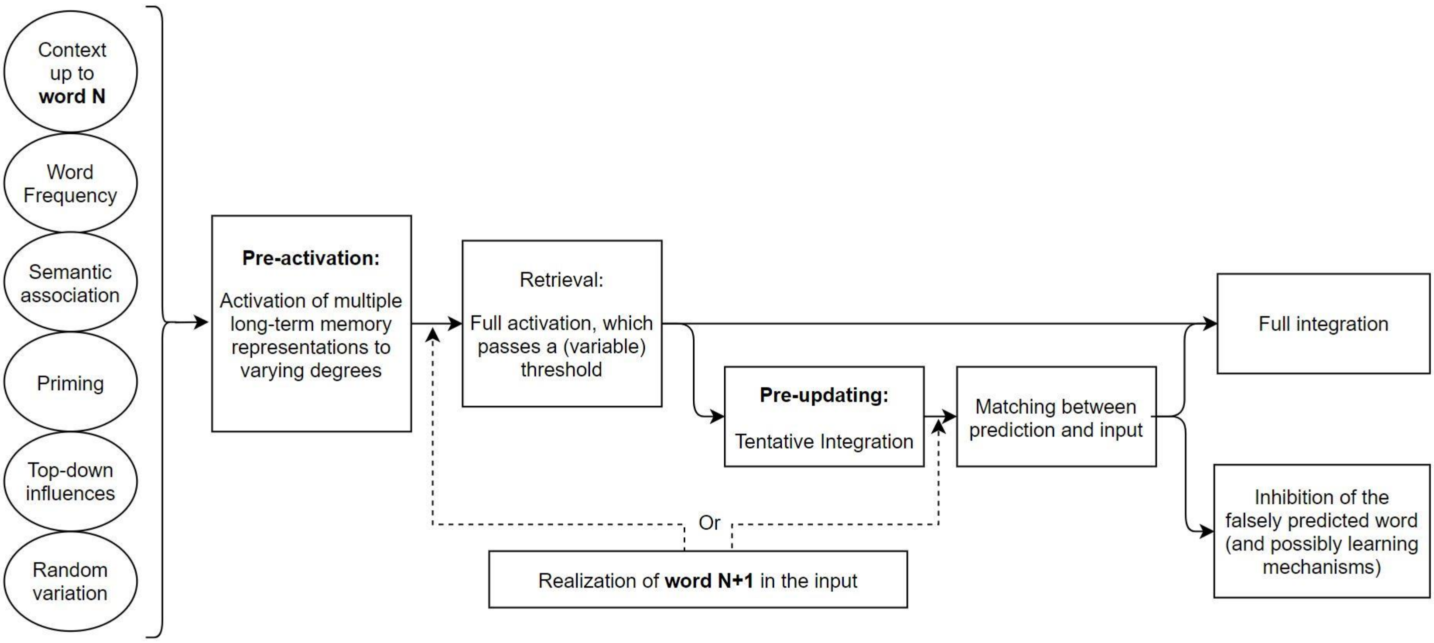

Figure 6: An outline of prediction mechanisms within the processing stages of a word

A question remains regarding the specificity of the predicted content that can be pre-updated. In the current study, the strongly constraining sentences led to a high likelihood that a specific word would appear. However, sentences can be constructed in a way that will lead to a high likelihood of occurrence not of a specific word, but of a word with a specific semantic or grammatical feature (e.g. animate, human, location, liquid substance, etc.; see Szewczyk & Schriefers, 2013). It is therefore possible that such features can also be pre-updated, even in the absence of prediction of a specific word. Whether or not this is plausible depends on how knowledge is stored in long-term memory, namely whether a feature can be highly activated when no specific word is highly activated, and on how representations are built in working memory, namely whether they can be partial and contain features of an upcoming word without the actual word. These questions closely relate to the suggestion by Kuperberg & Jaeger (2016), that pre-updating can occur at different representational levels. The P600 effects found in the current study provide a valuable tool for exploring this issue further.

5) References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247-264.

Boland, J. E. (2005). Visual arguments. *Cognition*, 95(3), 237-274.

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, 2(1).

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, 136, 135-149.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain research*, 1446, 127-143.

Chow, W. Y., & Phillips, C. (2013). No semantic illusions in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese. *Brain research*, 1506, 76-93.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why?. *Current directions in psychological science*, 19(1), 51-57.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, 19(4), 450-466.

Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, 48(9), 1203-1207.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117-1121.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2), 193-210.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20(6), 641-655.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of memory and Language*, 41(4), 469-495.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain research*, 1146, 75-84.

Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. *Brain and Language*, 87(3), 345-354.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, 47(2), 250-272.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology*, 33(4), 465-495.

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, 5(3), 475-494.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and language*, 140, 1-11.

Fruchter, J., Linzen, T., Westerlund, M., & Marantz, A. (2015). Lexical preactivation in basic linguistic phrases. *Journal of cognitive neuroscience*, 27(10), 1912-1935.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and cognitive processes*, 25(2), 149-188.

Green, C. (2017). Usage-based linguistics and the magic number four. *Cognitive Linguistics*, 28(2), 209-237.

Hagoort, P., Brown, C. M, & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and cognitive processes*, 8(4), 439-483.

Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In *The neurocognition of language* (pp. 273-317). Oxford University Press.

Hale, J. (2001, June). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies* (pp. 1-8). Association for Computational Linguistics.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32, 101–123.

Hale, J. (2016). Information-theoretical Complexity Metrics. *Language and Linguistics Compass*, 10(9), 397-412.

Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive brain research*, 19(1), 59-73.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and cognitive processes*, 15(2), 159-201.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, 49(1), 133-156.

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

Kim, A. E., Oines, L., & Miyake, A. (2018). Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(3), 406.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, cognition and neuroscience*, 31(1), 32-59.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and language*, 100(3), 223-237.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161–163.

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In: S. Kornblum, & J. Requin (Eds.), *Preparatory States and Processes* (pp. 217-237). Hillsdale, NJ: Lawrence Erlbaum.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, 25(3), 484-502.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.

Li, X., Zhang, Y., Xia, J., & Swaab, T. Y. (2017). Internal Mechanisms underlying Anticipatory Language Processing: Evidence from Event-Related-Potentials and Neural Oscillations. *Neuropsychologia*, 102, 70-81.

Linzen, T. (2009). Corpus of blog postings collected from the Israblog website. Tel Aviv: Tel Aviv University.

Linzen, T., & Jaeger, T. F. (2016). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*, 40, 1382–1411.

Lowder, M. W., Choi, W., Ferreira, F., & Henderson, J. M. (2018). Lexical predictability during natural reading: Effects of surprisal and entropy reduction. *Cognitive science*.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2), 81.

Ness, T., & Meltzer-Asscher, A. (in press). Lexical Inhibition due to Failed Prediction: Behavioral Evidence and ERP Correlates. *Journal of experimental psychology: Learning, memory, and cognition*.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, 22(12), 2886-2898.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, 31(6), 785-806.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 786.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22(3), 407-428.

Rommers, J., Dickson, D. S., Norton, J. J., Wlotko, E. W., & Federmeier, K. D. (2017). Alpha and theta band dynamics related to sentential constraint and word expectancy. *Language, cognition and neuroscience*, 32(5), 576-589.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(2), 344.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, 24(2), 232-252.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, 112(1), 1.

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, 68(4), 297-314.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302-319.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, 83(3), 382-392.

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1266.

Van Berkum, J. J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In *Semantics and pragmatics: From experiment to theory* (pp. 276-316). Palgrave Macmillan.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176-190.

Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, 8(3), 238-242.

Vos, S. H., & Friederici, A. D. (2003). Intersentential syntactic context effects on comprehension: the role of working memory. *Cognitive Brain Research*, 16(1), 111-122.

Wlotko, E. W., & Federmeier, K. D. (2007). Finding the right word: Hemispheric asymmetries in the use of sentence context information. *Neuropsychologia*, 45(13), 3001-3014.

Wu, S., Bachrach, A., Cardenas, C., & Schuler, W. (2010, July). Complexity metrics in an incremental right-corner parser. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 1189-1198). Association for Computational Linguistics.

2.2 Love thy neighbor: facilitation and competition between parallel predictions

Tal Ness and Aya Meltzer-Asscher (2021a)
Cognition, 207, 104509

Abstract

Ample evidence suggests that during word recognition and production, simultaneously activated lexical and sublexical representations interact, demonstrating varied patterns of facilitation and inhibition in various tasks and measures. A separate line of research has led to a growing consensus that prediction during sentence processing involves activating multiple possible predictions. However, very little is known about the nature of the interactions between parallel predictions. The current study employed a speeded cloze task to probe competition between simultaneously activated predictions. We focused on the modal response (the most probable completion for a sentence) and its strongest competitor (the second most probable completion). Examining production latencies of the modal response, the results showed an interaction between competitor strength and the semantic relatedness between the competitor and the modal: when the two were related, the stronger the competitor was, the more it facilitated production; however, when the two were unrelated, the stronger the competitor was, the more inhibition it caused. These results contrast with the pattern observed for the influences of near and distant semantic neighbors on word recognition and production. However, we show that when the different nature of the tasks is taken into consideration, these patterns of interaction between parallel predictions can be accounted for by the interactive activation and competition (IAC) model used to account for previous neighborhood effects (Chen & Mirman, 2012).

1) Introduction

One of the most basic processes necessary for language comprehension and production is lexical selection, namely, retrieval of a word from the mental lexicon while other words or concepts are also activated. Over the years, many studies investigated this process, leading to the general conclusion that the difficulty of lexical selection is highly dependent on how many other words are simultaneously activated, and to what degree. Notably however, although the influence of simultaneously activated words on lexical selection was demonstrated in numerous studies, the observed effects are not uniform, i.e. simultaneously activated words can either facilitate or inhibit retrieval. For example, many early studies showed reduced reaction times for words with many orthographic neighbors, relative to words with fewer neighbors, in tasks such as naming and lexical decision (e.g. Andrews, 1989, 1992; Forster & Shen, 1996; Johnson & Pugh, 1994; Sears, Hino, & Lupker, 1995). An orthographic neighbor is a word that differs from the target word by a single letter. When attempting to retrieve the target word, these neighbors are also activated due to their orthographic similarity to the target word. The studies above show that the more neighbors (of this type) are activated, the more they facilitate retrieval of the target word. However, orthographic neighbors were also shown to cause inhibition: when orthographic neighbors are more frequent than the target word, they inhibit its retrieval (e.g. Davis, Perea, & Acha, 2009; Ferraro & Hansen, 2002; Grainger, 1990; Grainger & Jacobs, 1996; Grainger, O'Regan, Jacobs, & Segui, 1989, 1992; Grainger & Segui, 1990). Additionally, transposed letter neighbors (i.e. words created from the target word by switching the positions of two adjacent letters) also inhibit retrieval (e.g. Acha & Perea, 2008; Andrews, 1996; Johnson, 2009). Similarly, phonological neighbors (i.e. words that are phonologically similar) can exert either facilitation or inhibition (see Dell & Gordon, 2003; Mirman, Kittredge, & Dell, 2010). Thus, although words that are activated due to similarity to the target word influence the difficulty of lexical selection, this influence may be facilitatory in certain circumstances and inhibitory in other.

1.1) Semantic neighborhood and semantic relatedness

Semantic neighbors are words that are related to the target word, sharing semantic features with it. Much like orthographic neighbors, semantic neighbors are activated with the target word due to their similarity, and therefore they also affect lexical retrieval. Some studies have shown that words with higher semantic neighborhood density, i.e. more semantic neighbors, are retrieved faster, both in recognition tasks (lexical decision and semantic decision) and production tasks (picture naming) (Buchanan, Westbury, & Burgess, 2001; Duñabeitia, Avilés, & Carreiras, 2008; Locker, Simpson, & Yates, 2003; Siakaluk, Buchanan, & Westbury, 2003; Yates, Locker, & Simpson, 2003). Importantly, however, later studies have shown that this facilitation is driven by distant semantic neighbors, i.e. neighbors that share few semantic features with the target word (Mirman, 2011; Mirman & Magnuson, 2008). Near semantic neighbors, on the other hand, cause the opposite effect, i.e. neighbors which share many semantic features with the target word cause inhibition, rather than facilitation (Fieder, Wartenburger, & Rahman, 2019; Mirman, 2011; Mirman & Magnuson, 2008).

These results suggest that high semantic relatedness between simultaneously activated words leads to competition between these words, hindering lexical retrieval. In apparent contradiction to this conclusion, several sentence processing studies suggest that this may not always be the case. The processing of an unexpected word within a sentence context was shown to be facilitated when it is semantically related to the predicted word (e.g. Brothers, Swaab, & Traxler, 2015; Federmeier & Kutas, 1999; Frisson, Harvey, & Staub, 2017; Luke & Christianson, 2016). For example, Federmeier & Kutas (1999) found that in a sentence such as 'The tourist in Holland stared in awe at the rows and rows of color. She wished she lived in a place where they grew …', facilitation was observed for the unexpected word 'roses' relative to the unexpected word 'palms' (reflected in this case in decreased amplitude of the N400 event-related potentials component), since 'roses' is more semantically related to the predictable word 'tulips'. Assuming that this facilitation stems from pre-activation of the semantically related predicted word, this suggests that semantic relatedness between simultaneously activated words can also cause facilitation. Note that the word 'tulips' is closely related to 'roses' and is thus comparable to a near semantic neighbor, shown to exert inhibition in recognition and naming studies. 'Tulip' is less related to 'palms' and is thus comparable to a distant semantic neighbor, shown to exert facilitation in these tasks. Nonetheless, the results of the sentence processing studies mentioned above are in the opposite direction, i.e. greater relatedness between the unexpected word and the predicted word causes facilitation. In line with these results, Roland, Yun, Koenig, and Mauner (2012) have also shown that reaction times in a self-paced plausibility judgement task (i.e. self-paced stops-making-sense task) were faster for a word the higher its average relatedness to alternative completions for the sentence was.

A crucial difference between the experiments in which close relatedness between activated words caused inhibition and the experiments in which it caused facilitation is that in the latter, the activations were induced by a sentence context. One of the main aims of the current study is therefore to further examine the interaction between simultaneously activated words during sentence processing, assessing the influence of semantic relatedness in this domain, and explaining why it may differ from typical semantic neighborhood effects in single-word tasks. Additionally, previous studies all examined the processing or production of a word that appeared in the input (either written, in isolation or in sentence context, or a picture of the word). In contrast, the current study focuses on interactions between parallel predictions induced by a sentence context that does not include the predicted words, manipulating the strength of the different predictions and the degree of relatedness between them.
Admittedly, alternative completions to the same sentence are always somewhat related to one another, being induced by the same context. However, sentences vary greatly in how semantically related their possible completions are. For example, the most probable completions for sentence (1) are 'popcorn' and 'candy', which are highly related and share many semantic features (e.g. edible items, snacks, treats, tasty, considered unhealthy, etc.). On the other hand, the most probable completions for sentence (2) are 'wheel' and 'mattress', which are not very related and do not share many semantic features other than the fact that they can both be inflated (in some cases). Thus, the related predictions 'popcorn' and 'candy', and the unrelated predictions 'wheel' and 'mattress', can be considered as the equivalents of near and distant semantic neighbors, respectively.

(1) Before the movie even started, the kids started to eat the ___
    C. Popcorn
    D. Candy
(2) Before the trip, Yoel looked for the pump in order to inflate the ___
    C. Wheel
    D. Mattress

In the current study we therefore test whether semantic relatedness between predictions during sentence processing elicits similar or opposed effects to the effects observed in single-word semantic neighborhood studies.

1.2) Parallel predictions and the speeded cloze task

Although, as discussed above, extensive research demonstrated influences of simultaneous activations on lexical selection, this phenomenon was hardly studied in regards to prediction during sentence processing. A recent study provides initial evidence for the influence of parallel predictions on one another. Staub and colleagues (2015) employed a speeded cloze task in order to assess the influence of cloze probability and sentence constraint on production onsets. Participants were presented with the beginning of a sentence (presented word by word at a fixed rate), and were instructed to produce a completion as quickly as possible. As in the common (non-speeded) cloze task, the cloze probability of each word was defined as the proportion of participants who produced this word as a completion for the sentence, reflecting how predictable the word is given the sentence; and sentence constraint was defined as the cloze probability of the sentence's most common response, reflecting how strong of a prediction the sentence encourages. The most probable completion of the sentences is termed the modal response.

Not surprisingly, the results showed that words with higher cloze probability (i.e. more predictable words) were produced fasted. This result indicates that the more the context activates a word, the faster it is retrieved and produced as a cloze response. More relevantly, the results also showed that words with low cloze probability are produced faster in high constraint versus low constraint sentences. Namely, a low cloze probability word is produced faster when the sentence has a highly probable alternative completion, compared to when it does not. As explained in Staub et al. (2015), these results suggest that multiple possible cloze responses are activated simultaneously, racing towards a retrieval threshold. Since the activation of each possible response, induced by the sentence context, is in correlation to its predictability (and therefore its cloze probability), the modal, most probable, response would most often reach the threshold first. However, due to random noise in the activation levels of each possible response, a less probable word can reach the threshold first. This means that when a low cloze word is produced even though a high cloze alternative is available, the activation of this low cloze word had to be exceptionally high in that moment (due to noise) relative to what is expected based on its probability. Otherwise, the high cloze word, which receives strong activation from the sentence, would have reached the threshold first, and the low cloze word would not have been produced. This is what gives rise to the influence of constraint on the onset of low cloze responses.

Notably, while Staub et al.'s (2015) finding that the strength of an alternative prediction can influence the onset of the ultimately produced response indicates that parallel predictions are simultaneously activated , these results can be explained without applying to direct influence of the activation of one response on the activation of another. Namely, the possible cloze responses can accumulate activation independently of one another, and the influence of constraint on the onset of a low cloze response merely stems from the fact that we can only measure production onset for a certain word when it is the first to reach retrieval threshold. A low cloze word has to be retrieved exceptionally fast in order for the modal, which has on average more activation, not to be produced. Thus, the influence of constraint on low cloze responses does not necessitate a direct interaction between the simultaneously activated predictions. Indeed, independent accumulation of activations is assumed in the model proposed by Staub et al. (2015) to capture these results (see section 4.1.2 for further consideration of this model).
Staub el al. (2015) have also tested whether the effect of constraint on production onset of low cloze responses, described above, can be attributed to semantic relatedness to the modal response. Potentially, the shorter production onset for low cloze responses in high constraint relative to low constraint contexts can be due to low cloze responses' relation to the highly probable word in the former. Looking at production onsets of non-modal responses, the authors found that although the effect of constraint was unlikely to be explained in full by semantic relatedness, semantic relatedness to the modal response did have an effect on the production onset of non-modal responses, such that words that were more related to the modal response were produced faster. This result is in line with the results observed in sentence processing studies (e.g. Brothers, Swaab, & Traxler, 2015; Federmeier & Kutas, 1999; Frisson, Harvey, & Staub, 2017; Luke & Christianson, 2016; Roland, Yun, Koenig, & Mauner, 2012), and in the opposite direction to semantic neighborhood effects observed in single-word tasks (Fieder, Wartenburger, & Rahman, 2019; Mirman, 2011; Mirman & Magnuson, 2008). Importantly, this effect of semantic relatedness in the speeded cloze task was only assessed with regards to production onset of non-modal responses. This means that in the trials from which these production onsets were taken, the activation levels within the participant's mind did not correspond to the cloze probability distribution (presumably due to random neural noise), since otherwise the modal response would 'win the race' towards the

retrieval threshold and the non-modal response would not have been produced. In order to provide an explanation that will reconcile the opposed influences of semantic relatedness in the different tasks, it is necessary to consider the underlying interactions between the simultaneously activated words, which is not possible when looking specifically at trials in which the underlying activations are atypical and unknown. In the current study we therefore focus on trials in which the modal response is produced, allowing us to provide additional insights by modeling the underlying activations.

1.3) The current study

The main aims of the current study were twofold: (i) to investigate the influence of relatedness between parallel prediction on lexical selection; and (ii) to find direct evidence for the influence of the activation of parallel predictions on one another. To achieve these goals, we employed a speeded cloze task, similar to Staub et al. (2015). We begin with an exploratory analysis of data from a previous experiment, followed up with a pre-registered confirmatory replication experiment. We additionally conduct an analysis as in Staub et al. (2015), aimed to provide an additional replication of their findings. Finally, we conduct simulations using Chan and Mirman's (2012) interactive activation and competition (IAC) computational model (see Discussion), adapting it to account for cloze response generation. We show that it is possible to account for the results of the current study, as well as the results of Staub et al. (2015), within the same model that accounts for previously observed neighborhood effects in single-word tasks.

In the current study, we focused on the production onset of the modal response, i.e. the most probable completion provided by participants in the cloze task. We asked whether the modal production onset is influenced by the strength (i.e. cloze probability) and the relatedness of its strongest competitor, i.e. the second most probable completion. If indeed alternative predictions are activated simultaneously, interacting with each other, then we should see an influence of the relatedness between the produced, modal word and its competitor. Since this influence stems from the activation of the competitor, it should be stronger the higher the competitor's cloze probability is. More specifically, if the influence of relatedness is similar to semantic neighborhood effects in single-word tasks, then simultaneously activated words should cause inhibition when they are highly related to the target word, and facilitation when they are remotely related. If this is the case, then when the modal and the competitor are highly related (i.e. the equivalent of near neighbors), the higher the competitor's cloze probability, the more inhibition it would cause, leading to increased production latencies for the modal word; when the modal and the competitor are unrelated (or remotely related, i.e. the equivalent of distant neighbors), on the other hand, the higher the competitor's cloze probability, the more facilitation it would cause, leading to decreased production latencies. However, if the influence of relatedness is in the opposite direction then that observed for semantic neighborhood effects, as exemplified in sentence processing studies (e.g. Brothers, Swaab, & Traxler, 2015; Federmeier & Kutas, 1999), then when the modal and competitor are highly related, the higher the competitor's cloze probability the more facilitation it would cause, leading to decreased production latencies of the modal word; conversely, for unrelated modals and competitors, the higher the competitor's cloze probability, the more inhibition it would cause, leading to increased production latencies.

2) Exploratory analysis of previous data

In a previous study (Ness & Meltzer-Asscher, submitted), a speeded cloze task was combined with EEG measurement, in order to study the pre-updating mechanism (Kuperberg & Jaeger; 2016; Lau, Holcomb,

& Kuperberg, 2013; Ness & Meltzer-Asscher, 2018). We re-analyzed the existing behavioral data from this study.

2.1) Materials and procedure

The materials and the data can be found at: https://osf.io/vjwds/?view_only=6758ee1a8b3f4e2b993b48a36cbb8afc. Forty-eight native Hebrew speakers participated in the experiment. The materials consisted of 156 Hebrew sentence beginnings, varying in constraint. The sentence fragment was presented word-by-word in the middle of the screen at a fixed rate (SOA = 600 ms), followed by a blank line prompting participants to produce a completion. Participants were instructed to provide the first completion that comes to mind, as quickly as possible. The sentences were composed in pairs such that each pair included a high constraint sentence and a low constraint sentence (based on a cloze probability questionnaire), in order to control for lexical material, but constraint was treated as a continuous variable in the current analysis (see Table 1 for example sentences). Presentation order was randomized for each participant. Sentences from the same set were separated by at least 50 trials.

Table 1: Example sentences

Constraint	Sentence frame
High	*Biglal še-ofir lo makir et ha-sifria, ha-safranit azra lo* ***limco*** *et* __ since that-ofir not know ACC the-library, the-librarian helped him **to-find** ACC __ 'Since Ofir isn't familiar with the library, the librarian helped him **find** __'
Low	*ofir xipes ve-xipes bemešex šaot, aval lo ecliax* ***limco*** *et* __ ofir searched and-searched for hours, but not succeeded **to-find** ACC __ 'Ofir had searched for hours, but he couldn't **find** __'

2.2) Analysis

For each item, we identified the most probable completion (the modal) and the second most probable completion (the competitor). We assessed the relatedness between these two possible completions for each item via a relatedness rating questionnaire, in which 30 participants (different from the ones participating in the main experiments) were asked to rate, on a 7-point scale, how related the two words were. Ratings were normalized within participant and the average relatedness rating was calculated for each word pair.

Production onset was marked using DeepWDM, a recurrent neural network for word duration measurement (Goldrick et al., 2018). A coder then listened to each of the recordings, transcribed the production, and corrected the marked onset when needed. Trials with speech errors, repairs or filled pauses were excluded (1.7% exclusion). Singular/plural and masculine/feminine forms of the same noun were collapsed by coding each noun as its singular-masculine form (e.g. 'shirts' and 'shirt' where counted as the same response). When the produced response consisted of more than one word, only the first word was coded unless the words formed a compound. Since the presented sentence fragments always ended with the Hebrew accusative case marker ‘et’, which only precedes definite nouns, the vast majority of produced responses began with a definite determiner (‘the’ - *ha*). Responses that included other determiners (e.g. *kol ha-kelim* – ‘all the-tools’) were coded without the determiner (such responses were very rare).

To test the interaction between parallel predictions, we fitted a linear mixed-effects model to the production onsets of modal responses (namely only for trials in which the most probable word was produced). The model included the fixed factors Constraint (equal to cloze probability of the modal) and the interaction between Competitor cloze (the cloze probability of the competitor) and Relatedness (the average rating of relatedness between the modal and the competitor). Main effects of these factors could not be included in the model due to their correlation with Constraint, resulting in multicollinearity (VIFs > 10). The model included random intercepts for participants. Random effects for items were not included since they could not be estimated independently of the fixed effects, as each item occurs at only one level of Constraint (as was done in Staub et al., 2015). Random slopes of both factors were initially included in the model, but had to be removed in order to achieve convergence.

2.3) Results

As expected, we found a significant effect of Constraint (Est. = -0.577, SE = 0.031, df = 3849, t = -18.47, $p < .001$), such that higher constraint led to shorter production onsets. Crucially, there was also a significant interaction between Competitor cloze and Relatedness (Est. = -0.325, SE = 0.066, df = 3849, t = -4.95, $p < .001$), such that when relatedness was high, higher cloze probability of the competitor led to shorter production onsets for the modal, but when relatedness was low, higher cloze probability of the competitor led to longer production onsets (See Figure 1). These results indicate that alternative predictions are activated simultaneously, interacting with each other. This is evidenced by the influence of the relatedness between the produced modal word and its competitor, which is greater the stronger the activation of the competitor is (i.e. the higher the competitor cloze is). Moreover, the results indicate that the influence of relatedness is such that a highly related competitor causes facilitation, while an unrelated competitor causes inhibition. This is in opposition to the results observed for near and distant semantic neighbors in single-word tasks.

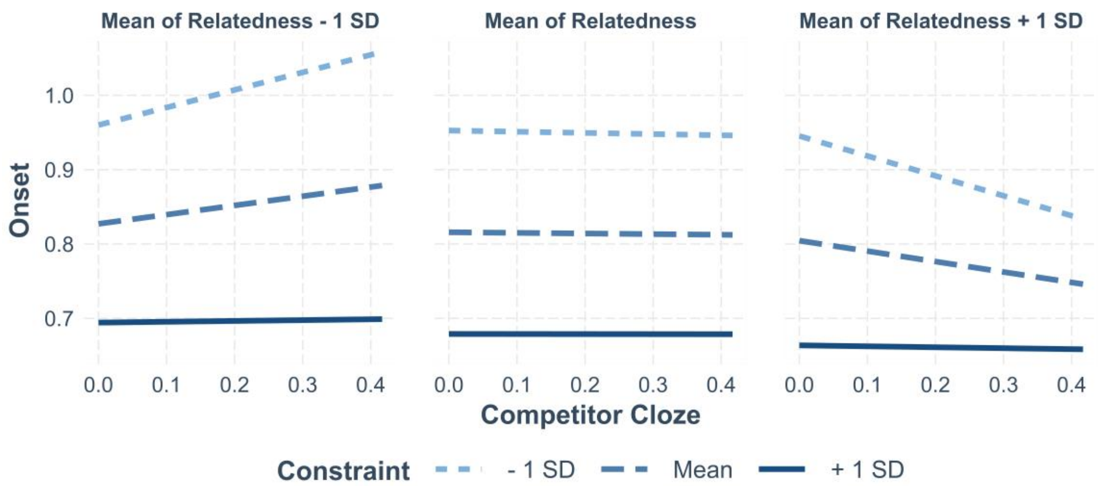

Figure 1. Onset times of modal responses in the data from Ness and Meltzer-Asscher (submitted), by cloze probability of the second most probable word ('Competitor cloze'), relatedness between the modal and the competitor, and sentence constraint. Regression lines are plotted.

In the analysis described above, the cloze probability of the competitor was used as a measure of competition strength. However, it is conceivable that a better measure would be the difference or the ratio between the cloze probabilities of the modal response and the competitor. We therefore created additional regression models, in which Competitor cloze was replaced by each of these measures.[5] The performance of the three variants of the model was compared in order to test which measure best captures the competition. The model with Competitor cloze outperformed the alternatives (Competitor cloze: AIC = 5106.5, BIC = 5137.9, Log likelihood = -2548.2; Ratio: AIC = 5122.2, BIC = 5153.6 , Log likelihood = -2556.1; Difference: AIC = 5123.4, BIC = 5154.8 , Log likelihood = -2556.7).

3) Replication experiment

Since the current hypotheses were developed after exploring the pre-existing data, we carried out an additional experiment, which was a replication study pre-registered on OSF, with the hypotheses and analysis pre-determined. The pre-registration report, as well as analysis code and data can be found at: https://osf.io/ab84y/?view_only=c6556beeb52d455e88067bfb5c2c610e, (the pre-registration report can be directly accessed at: https://osf.io/tzean).

3.1) Methods

3.1.1) Participants

The participants were 48 Tel-Aviv University students (16 males), all native Hebrew speakers, with an average age of 24.29 (range: 20-32). Participants were given course credit or were paid 40 NIS (~11$) for their participation. The experiment was approved by the Ethics Committee in Tel Aviv University.

3.1.2) Materials and procedure

The materials and procedure were identical to the previous experiment (see section 2.1 above), except that EEG was not recorded.

3.1.3) Audio recordings – transcription, onset measurement, and data analysis

Transcription, onset measurement, and data analysis of productions were done similarly to the previous experiment, except for the algorithm used for the initial automatic identification of production onset. Instead of DeepWDM, used in Ness & Meltzer-Asscher (submitted), in the current experiment we used a PRAAT script which identifies the onset of the first voiced segment in the production. Since the presented sentence fragments always ended with the Hebrew accusative case marker 'et', which only precedes definite noun phrases, the vast majority of produced responses began with 'ha-' ('the'), the definite determiner, which includes the voiced *a* vowel, enabling this script to provide an accurate onset for a large proportion of recordings compared to the previous experiment (the onset was manually corrected in 8.9% of the trials in the current experiment, as opposed to 17.5% in the previous experiment).

[5] In order to use comparable models with the same random effect structure, all achieving convergence, the random slopes had to be removed from these models.

Production onsets were analyzed with linear mixed-effects models. Analyses were conducted using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2014) in the R software environment (R Development Core Team, 2011). Data were winsorized by replacing data points exceeding 2.5 standard deviations (SD) from each participant's mean with the value of 2.5 SDs from that participant's mean (affecting 3.1% of the data). All models initially included the maximal random effects structure for participants (i.e. intercept and slopes of all fixed effects and interactions). The random effects structure was reduced when necessary in order to achieve convergence (the reduced models are specified where relevant), by iteratively removing the random slope associated with the smallest variance (Barr et al., 2013). Random effects for items were not included since they could not be estimated independently of the fixed effects, as each item occurs at only one level of Constraint (as was done in Staub et al., 2015).

3.2) Results

We first carried out the analysis reported in Staub et al. (2015) to examine the effects of constraint and cloze probability on production latencies. The results replicated Staub et al.'s (2015) findings, indicating that words with higher cloze probabilities are produced faster and that words with similar cloze probability are produced faster when the sentence's constraint is higher. This analysis is reported in Appendix A.

A relatedness rating questionnaire was conducted in order to assess the relatedness between word pairs (i.e. modal and competitor) that were not identical to those in the previous experiment, due to different participant completions in the two experiments. Thirty participants (different from the ones participating in the main experiments) were asked to rate, on a 7-point scale, how related the two words were. Ratings were normalized within participant and the average relatedness rating was calculated for each word pair.

As in the exploratory analysis reported above, to test for competition effects we fitted a linear mixed-effects model to the production onsets of modal responses (i.e. only for trials in which the most probable word was produced). The model included the fixed factors Constraint (equal to cloze probability of the modal) and the interaction between Competitor cloze and Relatedness. The random slope of Constraint (by participants) had to be removed in order to achieve convergence.

There was a significant effect of Constraint (Est. = -0.760, SE = 0.035, df = 3796, $t = -21.82$, $p < .001$), such that higher constraint led to shorter production onsets. Importantly, there was also a significant interaction between Competitor cloze and Relatedness (Est. = -0.212, SE = 0.088, df = 42.93, $t = -2.42$, $p = .02$), such that when relatedness was high, higher cloze probability of the second completion led to shorter production onsets, but when relatedness was low, higher cloze probability of the second completion led to longer production onsets (See Figures 2 and 3).

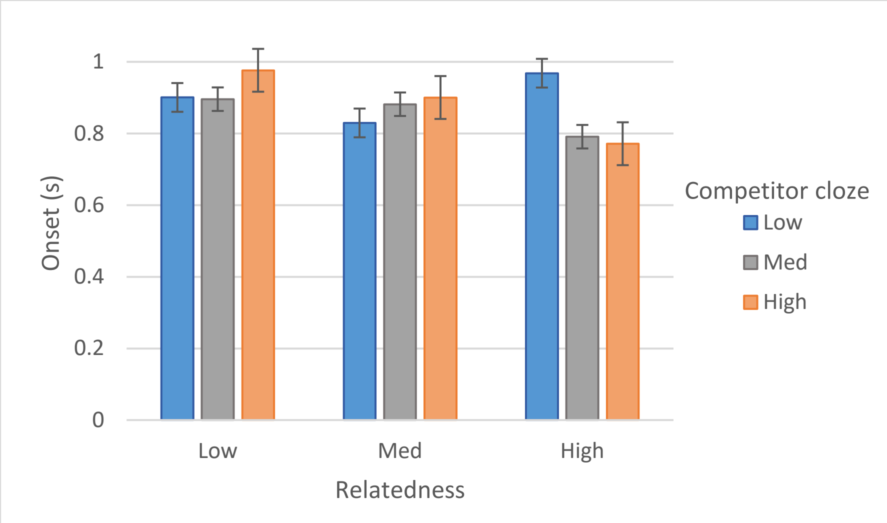

Figure 2. Average onset time of modal responses in the current experiment, by cloze probability of the second most probable word ('Competitor cloze'), and relatedness between the modal and the competitor. In order to divide the data into equal bins, the trials were first divided into three bins based on Relatedness percentiles (High/Medium/Low); then, the trials in each Relatedness category were divided into three bins based on Competitor cloze percentiles (High/Medium/Low).[6]

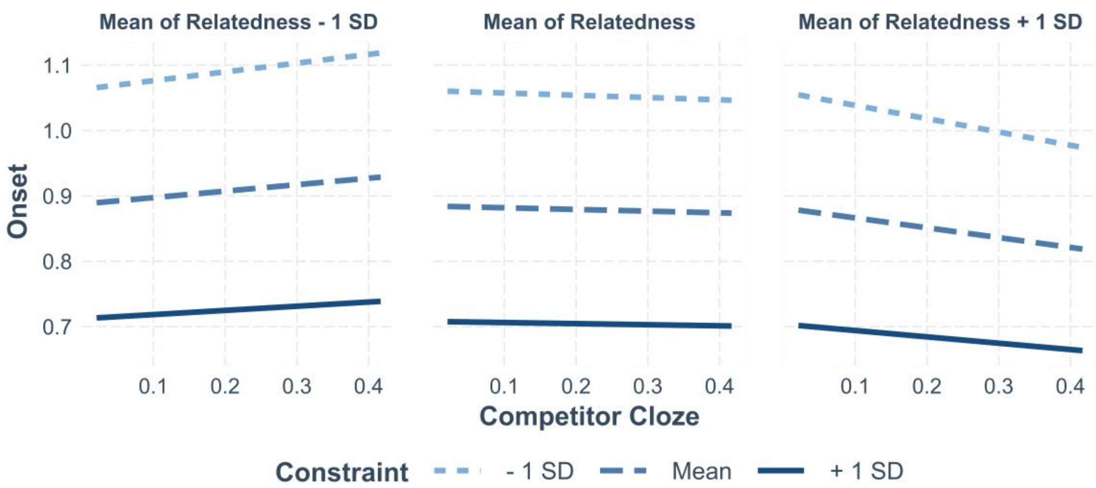

Figure 3. Onset time of modal responses in the current experiment, by cloze probability of the second most probable word ('Competitor cloze'), relatedness between the modal and the competitor, and sentence constraint. Regression lines are plotted.

[6] Due to correlation between Relatedness and Competitor cloze, dividing the trials into Relatedness categories and Competitor cloze categories independently of each other (i.e. setting fixed Second cloze boundaries across all Relatedness categories) was not possible, since it resulted in highly unequal numbers of trials in each bin.

We created additional regression models, in which Competitor cloze was replaced by either the difference or the ratio between the cloze probabilities of the modal response and the competitor.[7] Again, the model with Competitor cloze outperformed the alternatives (Competitor cloze: AIC = 5802.1, BIC = 5833.4 , Log likelihood = -2896.0; Ratio: AIC = 5805.8, BIC = 5837.1 , Log likelihood = -2897.9; Difference: AIC = 5809.7, BIC = 5841.0 , Log likelihood = -2899.8), confirming that Competitor cloze best accounts for the competition effects.

3.3) Discussion

The results of the pre-registered experiment replicate the competition effects observed in the exploratory analysis conducted with the behavioral data from Ness & Meltzer-Asscher (submitted). The results indicate that when the relatedness between the modal response and the second most probable response (the competitor) is high, higher cloze probability of the competitor leads to shorter production onsets for the modal response. However, when the relatedness between the modal response and the competitor is low, higher cloze probability of the competitor leads to longer production onsets for the modal response, providing evidence for competition between the (produced) modal word and the (not produced) second most probable word. These results indicate that alternative predictions are simultaneously activated, influencing the activations of one another.

The results also provide an additional replication of Staub and colleagues' (2015) findings regarding the influences of cloze probability and sentence constraint on production onset of cloze responses, showing that both high cloze probability and high constraint contribute independently to shorter production onsets.

4) Accounting for cloze response generation within the IAC framework

In two experiments, we found that activation of a closely related competitor facilitates retrieval of the modal response, while activation of an unrelated competitor inhibits retrieval of the modal response. These findings are in the opposite direction to the effects observed for near and distant semantic neighbors in single-word tasks. Namely, the activation of a closely related alternative prediction is facilitative, while the activation of a near neighbor is inhibitory. In this section we show that the current findings can be predicted by the same computational model that accounts for the effects of semantic neighbors, when the mechanisms underlying the different tasks are taken into consideration.

As discussed in the Introduction, accumulating evidence indicates an intriguing pattern of conflicting neighborhood effects under different circumstances, e.g. letter substitution neighbors cause facilitation (Andrews, 1989, 1992; Forster & Shen, 1996; Johnson & Pugh, 1994; Sears, Hino, & Lupker, 1995), while transposed letter neighbors cause inhibition (Acha & Perea, 2008; Andrews, 1996; Johnson, 2009); distant semantic neighbors cause facilitation (Mirman, 2011; Mirman & Magnuson, 2008), while distant semantic neighbors cause inhibition (Fieder, Wartenburger, & Rahman, 2019; Mirman, 2011; Mirman & Magnuson, 2008). To account for these contrasts, several models where put forward, describing the underlying interactions that lead to the observed variation in the effects. Most of these models are specifically designed to account for the results obtained in a certain domain, i.e. a specific task or modality (e.g. models accounting specifically for reading aloud include the Dual Route Cascaded

[7] In order to use comparable models with the same random effect structure, all achieving convergence, the random slopes had to be removed from these models.

Model, Coltheart, Davelaar, Jonasson, & Besner, 2001; the Connectionist Dual Process Model: Perry, Ziegler, & Zorzi, 2007; and the Self-Organizing Lexical Acquisition and Recognition Model: Davis, 2010). However, Chen and Mirman (2012) have constructed a general model, within the IAC framework (e.g., McClelland & Rumelhart, 1981), which is applicable in principle to any task or modality. Here, we adapt this model to account for the generation of predictions and cloze responses.

Chen and Mirman's (2012) model consists of three layers of processing units: a meaning layer in which each unit represents a semantic feature, a lexical layer in which each unit represents a word, and a word form layer in which each unit represents a phoneme or a letter. Each unit in the lexical layer is bidirectionally connected via facilitatory connections to the meaning units which represent its semantic features, as well as to the word form units which represent its constituent letters/phonemes. Additionally, each unit in the lexical layer (i.e. word) is connected to all other word units by bidirectional inhibitory connections. This lateral inhibition implements competition during lexical selection. Lateral connections within the meaning layer are either facilitatory or inhibitory, depending on the co-occurrence of the features, i.e. features that very often appear together exert facilitation on one another, while features that appear together rarely or not at all exert inhibition on one another. For simplicity, the model assumes that all words are connected to ten semantic features. Additionally, for simulations related to semantic neighborhood, the word form layer (phonemes or letters) was removed since it would not have any influence on the relevant effects. This layer would likewise not be relevant in the current simulations, and was removed from them.

Input activations enter the model either from the meaning layer or from the word form layer, depending on the simulated task. For example, in an auditory lexical decision task, the input activates the phonemes that the word consists of (which in turn activate units in the lexical layer). In a picture naming task, on the other hand, the input activates semantic features of the word (which in turn activate units in the lexical layer). In each iteration during the simulation, the amount of facilitation and inhibition that each unit receives from its connections is calculated, and activation levels are updated accordingly. Reaction times are defined as the number of iterations it took for a word unit to reach a pre-determined activation threshold, i.e. lexical selection is completed once a word reaches retrieval threshold.[8]
While the amount of facilitation exerted in the facilitatory connections between layers is linearly dependent on unit activation, the strength of inhibition exerted in the inhibitory connections within the lexical layer is a non-linear function of unit activation (namely, a sigmoid function), such that weakly activated word units exert very little inhibition, and strongly activated words exert very strong inhibition. Ultimately, whether a certain manipulation is predicted by the model to cause inhibition of facilitation depends on what is more significantly increased by the manipulation: the activation added to the target word via between-layers facilitatory connections, or the inhibition incurred to the target word via within-layer inhibitory connections.

The simulations conducted by Chen and Mirman (2012) have shown that the model correctly predicts the direction of neighborhood effects observed in a variety of tasks and modalities. Most relevantly, the model correctly predicts that in single-word tasks (e.g. word recognition and picture

[8] Here we only provide a description of aspects of the model that are relevant for our current simulations. For a full description of the model and previous simulations see Chen and Mirman (2012).

naming) many near semantic neighbors (relative to few) cause inhibition, while many distant neighbors (relative to few) cause facilitation. Near semantic neighbors are modeled as words that have high overlap in semantic features with the target word, i.e. connected to 8 (out of 10) of the target word's semantic features. In this case, the neighbor will be almost as strongly activated by the target's semantic features as the target itself. Thus, the activation of the neighbor at the lexical level would be strong and the inhibition it would exert on the target word would outweigh the additional facilitation that the target word would get from the shared features (due to their facilitatory connections with the neighbor being bidirectional). This would be the result mainly since, as mentioned above, the inhibition is non-linearly dependent on unit activation (while the facilitation is linearly dependent on unit activation) and would thus be greater than the facilitation when the neighbor's word unit activation is high. On the other hand, distant semantic neighbors, which are modeled as words that are connected to 4 (out of 10) of the target word's semantic features, are weakly activated by the target's semantic features. Therefore, the activation of the neighbor at the lexical level would be weak, exerting very little inhibition on the target word, which would be outweighed by the additional facilitation that the target word would get from the shared features (due to their bidirectional facilitatory connections with the neighbor).

4.1.1) Relatedness and competitor cloze

Crucially, in Chen and Mirman's (2012) simulations, the input is assumed to only activate the features of the target word, and other words are activated due to their connections with shared features. This assumption is appropriate for e.g. word recognition and picture naming, for which the input is the word itself, or a picture of it. However, this is not the case when considering cloze responses, for which the input is a sentence context. A given context can generate several alternative predictions, activating the features of various words to different degrees based on their probabilities. This means that words other than the produced cloze response (which we will refer to as the target word henceforth) receive activation directly from the input, and not only by association to the target word.

As mentioned in the Introduction, sentences differ in the degree to which alternative completions share semantic features with each other. Returning to the example discussed in the introduction, the most probable completions for sentence (1) are 'popcorn' and 'candy', which share many semantic features, while the most probable completions for sentence (2) are 'wheel' and 'mattress', which barely share any semantic features. Therefore, we assume that in sentence (2), the activation of the features of the target word 'wheel' is proportional to the probability of this word only. In sentence (1), the activation of the features which are unique to the target word 'popcorn' is also in proportion to this word's cloze probability; however, we assume that the features shared by 'popcorn' and 'candy' are activated in proportion to the sum of the cloze probabilities of these two words. The input activations of the target word are thus not based only on the target word's cloze probability, but also on the cloze probability of alternative predictions with shared semantic features. In other words, the target word can benefit from input activation of the competitor if it is semantically related.

Indeed, with the additional assumption that in the cloze task the input activation of each feature is proportional to the sum of the cloze probabilities of the cloze responses that share it, the model predicts the results observed in the current study[9]. In the first simulation we modeled a sentence in which the

[9] The code for our simulations can be found at: https://osf.io/ab84y/?view_only=c6556beeb52d455e88067bfb5c2c610e

constraint (i.e. the cloze probability of the modal/target word) is 0.6, and the competitor's cloze is 0.4, namely, a strong competitor. We manipulated the relatedness between the target and the competitor such that a related word shared 8 (out of 10) semantic features with the target word, and an unrelated word shared only one feature.[10] The results of the simulation are shown in Figure 4A. These results show faster activation and decreased reaction time for the target when the competitor is related to the target word, relative to when it is unrelated, in line with the results observed in the current study.

In an additional simulation, we explored the influence of the competitor's cloze on the relatedness effect. This simulation was identical to the first one, except that the competitor's cloze was 0.2 (instead of 0.4 in the first simulation). The results of the simulation are shown in Figure 4B. These results show a smaller facilitation effect by a related competitor, and a smaller inhibition effect by an unrelated competitor, relative to the first simulation.

The two simulations together thus show an interaction between competitor cloze and relatedness, as observed in the current study. For a closely related competitor, the higher its cloze probability, the more it causes facilitation, and for an unrelated competitor, the higher its cloze probability, the more it causes inhibition. Thus, although the current results contrast with the results observed for semantic neighborhood effects, the opposite direction of the effects stems from the characteristics of the input activations, and can be predicted within the same computational model.

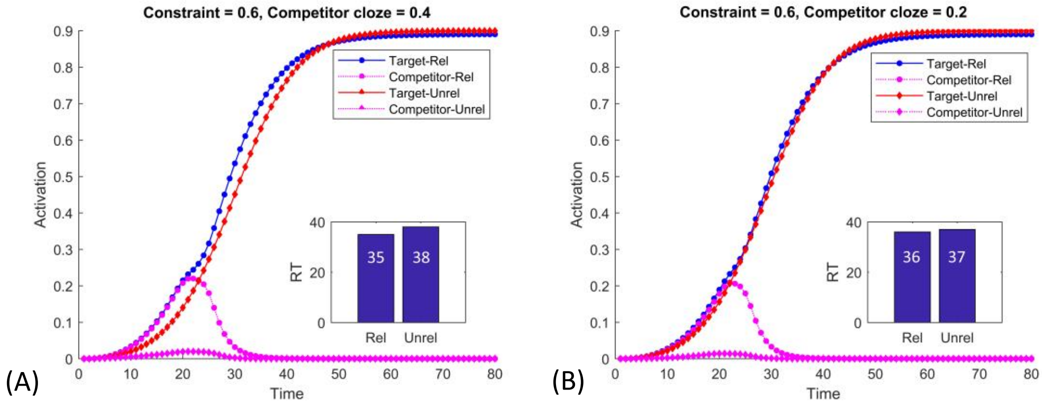

Figure 4. Simulation results: activation levels over time for the target word (i.e. the produced modal response) and the competitor, when the target and the competitor are related (Rel) or unrelated (Unrel). The competitor's cloze probability is 0.4 in panel (A) and 0.2 in panel (B). The bar plots show reaction times (RT) predicted by the model, i.e. the number of iterations until the target word reached the retrieval threshold of 0.7.

4.1.2) The influence of constraint on the production of low cloze responses

As discussed in the Introduction, Staub and colleagues (2015) have shown that words with higher cloze probability are produced faster in the speeded cloze task, and that words with similar (low) cloze

[10] We note that modeling the unrelated word as not sharing any features with the target word, rather than sharing one feature, did not significantly alter the results, and the direction of effects predicted by the model remained the same. We chose to model the 'unrelated' word as sharing a (single) feature with the target word in order to acknowledge that words that are activated by the same sentence context cannot truly be completely unrelated (otherwise there would not be a context that could be completed by both).

probabilities are produced faster in high constraint (relative to low constraint) sentences. These results were replicated in Ness & Meltzer-Asscher (submitted) as well as in the current experiment. In their paper, Staub and colleagues (2015) suggested a computational model to account for these results. However, in that model each potential response independently accumulates activation, and the activation of one response cannot influence the activation of another. Thus, while the model accounts for the influences of cloze probability and constraint, it cannot (in its current form) account for any influence of relatedness between alternative responses.

We therefore sought to adapt Chan and Mirman's (2012) IAC model to fully capture the generation of cloze responses, including the influences of cloze probability and sentence constraint. We did this by the addition of random noise. In the current model, without noise, the modal response will always be the first to reach the threshold. However, participants do not uniformly produce the most probable word in the cloze task, which indicates that the race towards the retrieval threshold is noisy, allowing for a less probable word to occasionally win over a more probable one. The addition of noise to the model was implemented by adding a random number, drawn from a normal distribution with a mean of zero, to the activation level of each unit in each iteration. The random noise could therefore either increase or decrease the activation level.

We then conducted two additional simulations, in order to test the influences of cloze probability and sentence constraint. In the first simulation we modeled a sentence for which the possible cloze responses have cloze probabilities of 0.6 and 0.4 (i.e. the sentence constraint is 0.6). In the second simulation we modeled a sentence for which the possible cloze responses are 0.4, 0.2, 0.2, 0.2 (i.e. the sentence constraint is 0.4). For simplicity, none of the cloze responses in these simulations were related (i.e. each word was connected to 10 different semantic features). Each simulation was run 10000 times. The results are plotted in figure 5. First, higher cloze probability led to reduced reaction times (see figure 5A), since the higher the cloze probability of a word is, the higher the input activation it receives. Additionally, looking at reaction times for the words with 0.4 cloze probability in both simulations (namely when sentence constraint is 0.6 vs. when it is 0.4), reaction times were lower in the high constraint sentence then in the low constraint sentence (see figures 5B and 5C). This is in line with the explanation discussed in the introduction (proposed by Staub et al., 2015), that reaction times for low cloze responses in high constraint sentences are decreased simply since the high cloze alternative obscures the measurement of potentially longer retrieval times. Namely, if the low cloze word had not been retrieved fast enough to win over the high cloze alternative, it would not have been produced, since the high cloze alternative would be produced. The results of the simulations therefore capture the behavioral findings regarding the influences of cloze probability and sentence constraint on production onsets.

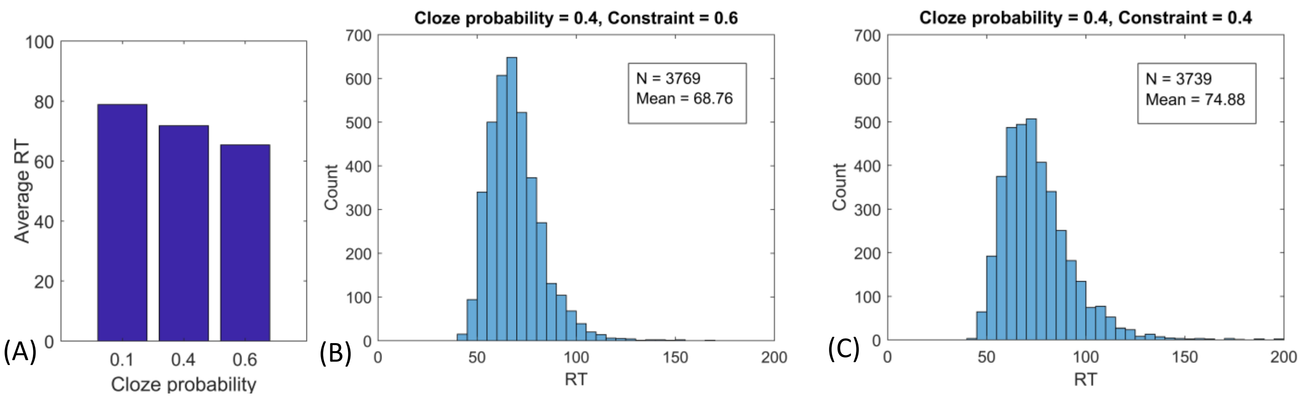

Figure 5. Simulation results: average reaction time by cloze probability (A). Reaction times for the 0.4 cloze probability response in high (A) and low (B) constraint.

5) General discussion

The current study employed a speeded cloze task in order to examine the interaction between parallel predictions. We assessed whether, and how, production onsets of the modal response are influenced by the strength and relatedness of its strongest competitor, the second most probable completion. In an exploratory analysis using previous data, as well as in a pre-registered replication, we found that when the modal response and the competitor were related to each other, higher cloze probability of the competitor led to shorter production onsets of the modal response. This means that activation of the related competitor facilitated the retrieval of the modal response. However, when the competitor was unrelated to the modal response, higher cloze probability of the competitor led to longer production onsets of the modal response. This means that activation of the unrelated competitor inhibited the retrieval of the modal response. These results provide direct evidence for the prevalent assumption that multiple predictions are activated simultaneously. More specifically, the results demonstrate interaction between parallel predictions, showing for the first time that parallel predictions do not accumulate activations independently, but that instead, the activation levels of different predictions directly influence one another.

Additionally, the results show that the influence of relatedness between parallel predictions during sentence processing is in the opposite direction than the effects observed for near and far semantic neighbors in single-word tasks. Namely, the activation of a closely related alternative prediction in the current study is facilitative, while the activation of a closely related neighbor was found in previous studies to be inhibitory. We show that although these effects of semantic relatedness manifest in different directions, they both stem from dynamic interactions between and within levels of representation. The current results can thus be explained by taking into consideration the different inputs to lexical selection in prediction during sentence processing and cloze response generation, as compared to single-word tasks such as word recognition or picture naming. Specifically, since in single-word tasks the input is the word itself (or a picture of it), this input can be assumed to only activate the features of the target word, and other words are activated only due to their connections with shared features. With a sentence context, on the other hand, a given context generates several alternative predictions, directly activating the features of various words to different degrees based on their probabilities. This means that words other than the target word receive activation directly from the input, and thus, the target word can benefit from input activation of the competitor if it is semantically related.

The fact that the same model accounts for behavior in both word recognition and cloze response generation despite the different manifestation of relatedness effects indicates that the generation of predictions occurs under the same principles that operate during the retrieval of a word when it is presented in the input. The results and simulations thus hint that the underlying architecture that supports prediction is the same as that underlying word processing in general.

Previous results and model simulations have led Chen and Mirman (2012, 2015) to suggest that, in general, simultaneously activated words will exert net inhibitory effects if they are strongly activated, and net facilitative effects if they are weakly activated. This was most directly demonstrated by showing

that the same phonological neighbors exert inhibition when they are strongly activated, and facilitation when their activation is reduced (due to inhibition from activated semantic competitors; Chen & Mirman, 2015). Notably, the current study indicates that this generalization does not hold in all instances. Our results show that stronger absolute activation of a competitor can be facilitatory if this activation adds to the target word's activation as well (i.e. a related alternative prediction). As the target word has higher activation levels relative to the competitor, the addition of activation to shared features contributes more to the target word then to the competitor due to the non-linear inhibition at the lexical level, which means that the stronger the related competitor is, the more it would facilitate the retrieval of the target word.
As explained above, our account for the results emerges from Chen and Mirman's (2012) model, and we attribute the observed effects to the balance between activations from semantic features and inhibition at the lexical level. While this model provides a plausible mechanism that captures our results (as well as previous results), we would like to acknowledge that there could be alternative architectures that can potentially account for the same pattern. For example, to capture the current data we can consider an architecture that contains a level of event representations, or situation models.[11] A sentence fragment for which the two most probable completions are highly related likely represents a less ambiguous situation than a sentence fragment for which the two most probable completions are unrelated. For example, looking again at the sentences in (1)-(2) discussed above, a sentence such as 'Before the movie even started, the kids started to eat the __' represents very similar scenarios whether it is completed by 'popcorn' or by 'candy'. In this case, a strong competitor that is highly related to the modal response indicates that there is little ambiguity in the situation model that the sentence fragment evokes, and therefore responses are fast. On the other hand, a sentence such as 'Before the trip, Yoel looked for the pump in order to inflate the __' can represent somewhat different scenarios when it is completed by 'wheel' and when it is completed by 'mattress' (e.g. a planned bike ride versus a planned camping trip). In this case, a strong competitor that is not highly related to the modal response indicates that there is substantial ambiguity in the situation model that the sentence fragment evokes, which may delay responses. Thus, the current results may also be explained as competition and facilitation at the level of event representation. The current study cannot decide between the two accounts, although we do note that the account based on Chen and Mirman's (2012) model has the advantage of adding very minimal assumptions to a model that already accounts for a wide range of phenomena. Further research is needed in order to tease apart representations at the lexical level from those at the situation level.

The study reported here can form a basis for several future directions. First, one limitation of the current study is that in investigating the interaction between parallel predictions, we specifically tested the influence of the strongest competitor, namely the second most probable word, on production of the modal word. However, it is likely that the entire distribution of possible responses influences the production of any possible response. Looking at these complex effects posits some interesting methodological and theoretical considerations, such as how to create measures that weigh the influences of multiple words, taking into account not only the probability of each response but also the degree to which each word is related to the produced word, and to the other alternatives.
Additionally, the current study focused on the effect of semantic relatedness between predictions. However, neighborhood effects show that orthographic and phonological similarity between simultaneously activated words can also influence lexical selection. Further research is needed in order

[11] We thank Adrian Staub for proposing this alternative account.

to test whether orthographic and phonological similarity between parallel predictions has an effect on cloze response generation, and in what direction.

Lastly, the influence of parallel predictions on one another can potentially take place at different stages – during activation, selection/retrieval, maintenance in working memory, production planning and so on. We have shown that competitor cloze, rather than the difference or ratio between cloze probabilities of the modal response and the competitor, best accounted for the interaction we observed. This finding suggests that the interaction found here likely stems from competition at the activation stage, rather than difficulty in selection mechanisms and/or retrieval, since difficulty stemming from competition in these latter stages is likely to depend on the modal response's strength over the alternative rather than directly on how strong the alternative is in absolute terms. Moreover, the model used here (Chen & Mirman, 2012), which assumes interaction during the activation of words (prior to retrieval), was able to capture the behavioral results. However, the interaction between parallel predictions may not be limited to this stage only. Further research can pinpoint the specific stage, or stages, that are relevant for these interactions, in order to better understand the interplay between simultaneously activated words.

6) References

Acha, J., & Perea, M. (2008). The effect of neighborhood frequency in reading: Evidence with transposed-letter neighbors. Cognition, 108, 290–300. https://doi.org/10.1016/j.cognition.2008.02.006

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search?. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(5), 802. https://doi.org/10.1037/0278-7393.15.5.802

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. Journal of Memory and Language, 35, 775–800. https://doi.org/10.1006/jmla.1996.0040

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 234–254. https://doi.org/10.1037/0278-7393.18.2.234

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. https://doi.org/10.1016/j.jml.2012.11.001

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. Cognition, 136, 135-149. https://doi.org/10.1016/j.cognition.2014.10.017

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. Psychonomic Bulletin & Review, 8, 531–544. https://doi.org/10.3758/BF03196189

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: toward a unified account of facilitative and inhibitory effects of lexical neighbors. Psychological review, 119(2), 417. https://doi.org/10.1037/a0027175

Chen, Q., & Mirman, D. (2015). Interaction between phonological and semantic representations: Time matters. Cognitive Science, 39(3), 538-558.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychological Review, 108, 204–256.

Davis, C. J. (2010). The spatial coding model of visual word identification. Psychological Review, 117, 713–758. https://doi.org/10.1037/a0019738

Davis, C. J., Perea, M., & Acha, J. (2009). Re(de)fining the orthographic neighborhood: The role of addition and deletion neighbors in lexical decision and reading. Journal of Experimental Psychology: Human Perception and Performance, 35, 1550–1570. https://doi.org/10.1037/a0014253

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes. In N.O. Schiller, A.S. Meyer (Eds.), Phonetics and phonology in language comprehension and production: Differences and similarities, 6, 9-37. Berlin, Germany: Mouton de Gruyter.

Dun˜abeitia, J. A., Avile´s, A., & Carreiras, M. (2008). NoA's ark: Influence of the number of associates in visual word recognition. Psychonomic Bulletin & Review, 15, 1072–1077. https://doi.org/10.3758/PBR.15.6.1072

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. Journal of memory and Language, 41(4), 469-495. https://doi.org/10.1006/jmla.1999.2660

Ferraro, F. R., & Hansen, C. L. (2002). Orthographic neighborhood size, number of word meanings, and number of higher frequency neighbors. Brain and Language, 82, 200–205. https://doi.org/10.1016/S0093-934X(02)00016-0

Fieder, N., Wartenburger, I., & Rahman, R. A. (2019). A close call: Interference from semantic neighbourhood density and similarity in language production. Memory & cognition, 47(1), 145-168. https://doi.org/10.3758/s13421-018-0856-y

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 696–713. https://doi.org/10.1037/0278-7393.22.3.696

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. Journal of Memory and Language, 95, 200-214.

Goldrick, M., McClain, R., Cibelli, E., Adi, Y., Gustafson, E., Moers, C., & Keshet, J. (2018). The influence of lexical selection disruptions on articulation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 45(6), 1107-1141. https://doi.org/10.1037/xlm0000633

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. Journal of Memory and Language, 29, 228–244. https://doi.org/10.1016/0749-596X(90)90074-A

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. Psychological Review, 103, 518–565. https://doi.org/10.1037/0033-295X.103.3.518

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. Perception & Psychophysics, 45, 189 –195. https://doi.org/10.3758/BF03210696

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. Perception & Psychophysics, 51, 49–56. https://doi.org/10.3758/BF03205073

Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. Perception & Psychophysics, 47, 191–198. https://doi.org/10.3758/BF03205983

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. Language, cognition and neuroscience, 31(1), 32-59. https://doi.org/10.1080/23273798.2015.1102299

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models. R Package Version 2.0–11. Available at: https://cran.r-project.org/package=lmerTest

Johnson, R. L. (2009). The quiet clam is quite calm: Transposed-letter neighborhood effects on eye movements during reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35, 943–969. https://doi.org/10.1037/a0015572

Johnson, N. F., & Pugh, K. R. (1994). A cohort model of visual word recognition. Cognitive Psychology, 26, 240 –346. https://doi.org/10.1006/cogp.1994.1008

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. Journal of Cognitive Neuroscience, 25(3), 484-502. https://doi.org/10.1162/jocn_a_00328

Locker, L., Simpson, G. B., & Yates, M. (2003). Semantic neighborhood effects on the recognition of ambiguous words. Memory & Cognition, 31, 505–515. https://doi.org/10.3758/BF03196092

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. Cognitive Psychology, 88, 22-60.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. Psychological Review, 88, 375–407. https://doi.org/10.1037/0033-295X.88.5.375

Mirman, D. (2011). Effects of near and distant semantic neighbors on word production. Cognitive, Affective, and Behavioral Neuroscience, 11, 32–43. https://doi.org/10.3758/s13415-010-0009-7

Mirman, D., Kittredge, A. K., & Dell, G. S. (2010). Effects of near and distant phonological neighbors on picture naming. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 32, No. 32). https://escholarship.org/uc/item/5620c08n

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34, 65–79. https://doi.org/10.1037/0278-7393.34.1.65

Ness, T. & Meltzer-Asscher, A. (2018). Predictive pre-updating and working memory capacity: Evidence from event-related potentials. Journal of Cognitive Neuroscience, 30(12), 1916-1938. https://doi.org/10.1162/jocn_a_01322

Ness, T. & Meltzer-Asscher, A. (submitted). From pre-activation to pre-updating: A threshold mechanism for commitment to strong predictions.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. Psychological Review, 114, 273–315. https://doi.org/10.1037/0033-295X.114.2.273

Roland, D., Yun, H., Koenig, J. P., & Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. Cognition, 122(3), 267-279.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 21, 876–900. https://doi.org/10.1037/0096-1523.21.4.876

Siakaluk, P. D., Buchanan, L., & Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. Memory & Cognition, 31, 100–113. https://doi.org/10.3758/BF03196086

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. Journal of Memory and Language, 82, 1-17. https://doi.org/10.1016/j.jml.2015.02.004

Yates, M., Locker, L., & Simpson, G. B. (2003). Semantic and phonological influences on the processing of words and pseudohomophones. Memory & Cognition, 31, 856–866. https://doi.org/10.3758/BF03196440

7) Appendix A – replication of Staub et al. (2015)'s analysis with the current data

We followed the analysis reported by Staub et al. (2015). Model 1 included cloze probability as the only fixed effect. In order to achieve convergence, the random slopes had to be removed in this model. Higher cloze probability led to shortened production onset (Est. = -0.279, SE = 0.007, t = -39.35, $p < .001$). Model 2 included constraint as a fixed factor, as well as a factor indicating whether or not the modal response was produced. In order to achieve convergence, the random slopes of Constraint and the interaction between Constraint and Modal had to be removed in this model. A significant effect of Constraint was found (Est. = -0.192, SE = 0.008, t = -22.37, $p < .001$), such that higher constraint led to shortened production onset. Additionally, modal responses were produced significantly faster than non-modal responses (Est. = -0.247, SE = 0.030, t = -8.20, $p < .001$), indicating that the effect of cloze probability in Model 1 cannot be solely driven by constraint (See Figure 6A). Model 3 included cloze probability and constraint as fixed factors, and only included responses with cloze probability <.4, allowing to separate constraint and cloze probability. In order to achieve convergence, all random slopes had to be removed in this model. Significant effects of both cloze probability and constraint were found (cloze probability: Est. = -0.167, SE = 0.011, t = -15.20, $p < .001$; constraint: Est. = -0.135, SE = 0.012, t = -10.97, $p < .001$), confirming that the production of low cloze completions is indeed influenced by sentence constraint (See Figure 6B). The results of the current experiment therefore provide an additional replication of the findings of Staub and colleagues (2015). The full results of the three models are reported in Table 2.

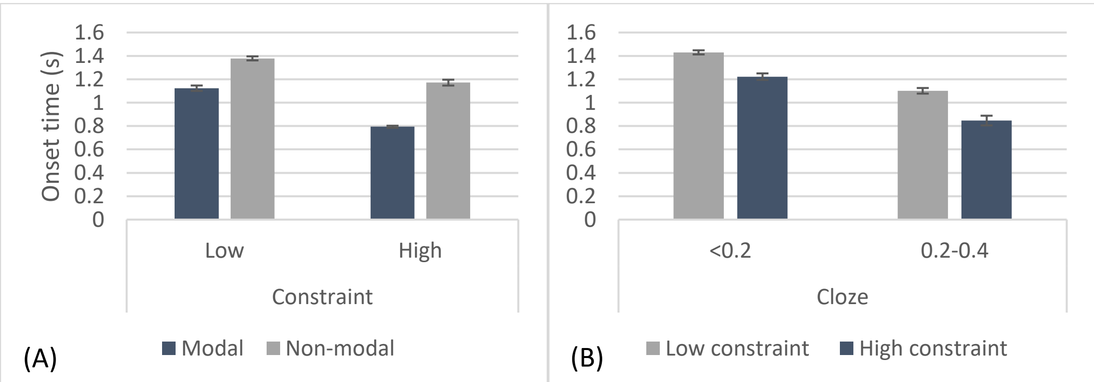

Figure 6. (A) Mean onset time for modal and non-modal responses, in sentences with high (>0.5) and low (<=0.5) constraint. (B) Mean onset time by cloze probability, in sentences with high (>0.5) and low (<=0.5) constraint. Error bars represent standard error of the mean.

Table 2: Mixed-effects regression model coefficients

	Estimate	SE	df	t value	p value
Model 1					
Cloze probability	0.279	0.007	7316	-39.35	< .001
Model 2					
Constraint	0.192	0.008	7302	-22.37	< .001
Modal	-0.247	0.030	58.76	-8.20	< .001
Constraint : Modal	-0.020	0.017	7295	-1.18	0.240
Model 3					
Cloze probability	-0.167	0.011	3910	-15.20	< .001
Constraint	-0.135	0.012	3907	-10.97	< .001
Cloze probability : Constraint	-0.025	0.016	3907	-1.56	0.118

2.3 From pre-activation to pre-updating: A threshold mechanism for commitment to strong predictions

Tal Ness and Aya Meltzer-Asscher (2021b)

Psychophysiology, 58(5), e13797

Abstract

Ample evidence suggests that during sentence processing comprehenders can "pre-activate" lexical/semantic knowledge stored in long-term memory. A relatively recent development suggests that in some cases a stronger form of prediction is employed, involving "pre-updating" the predicted content into the sentence's representation being built in working memory. The current study argues for an activation threshold mechanism by which pre-updating is initiated, within the routine processing stages of a word in a context. By combining a speeded cloze task with event-related potentials, we were able to analyze electrophysiological data measured prior to when participants were prompted to produce a completion, based on the participant's cloze response, reflecting their strongest prediction at that specific moment in time. A P600 effect reflecting pre-updating was observed in high (relative to low) constraint sentences, even in trials where the participant predicted a low cloze word. The results support a mechanism in which multiple predictions accumulate activations, 'racing' towards a retrieval threshold. Once the activation level of a certain word passes the threshold, the word is integrated into the sentence representation in working memory. Pre-updating occurs if a certain prediction passes retrieval threshold prior to its realization in the input.

1) Introduction

With the accumulation of a large body of evidence indicating that prediction plays a role in sentence processing, numerous recent studies have begun to explore the nature and dynamics of prediction, in order to gain a fine-grained understanding of its underlying mechanisms (e.g. Brothers, Swaab, & Traxler, 2015; DeLong, Quante, & Kutas, 2014; Ito et al., 2016; Kuperberg, Brothers, & Wlotko, 2020; Lowder et al., 2018; Staub, 2011; Szewczyk & Wodniecka, 2020; Wlotko & Federmeier, 2015; see Federmeier, 2007; Ferreira & Chantavarin, 2018; Kuperberg & Jaeger, 2016; Van Petten & Luka, 2012, for relevant reviews).

Prediction in sentence processing is frequently suggested to manifest as "pre-activation" of lexical/conceptual knowledge stored in long-term memory. The activation level of predicted content can increase due to spreading activation from previous linguistic material, or due to more controlled prediction processes, facilitating its retrieval when it appears in the input. Pre-activation was suggested to underlie a wide range of findings showing reduced processing difficulty for predictable as compared to unpredictable words, demonstrated in reduced reaction times (e.g. Ehrlich & Rayner, 1981; Schwanenflugel & Shoben, 1985; Traxler & Foss, 2000), decreased amplitudes of the N400 event-related potentials (ERP) component (e.g. Delong, Urbach & Kutas, 2005; Kutas & Hillyard, 1984; Wlotko & Federmeier, 2012) and other measures.

A more recent development suggests that in some cases, a stronger form of prediction can occur. When a certain prediction is highly activated, it can be "pre-updated", i.e. integrated into the sentence's representation being built in working memory (Kuperberg & Jaeger, 2016; Lau, Holcomb, & Kuperberg, 2013; Ness & Meltzer-Asscher, 2018a). This distinction between pre-activation and pre-updating highlights the difference between priming of multiple entities (pre-activation) and commitment to a specific prediction (pre-updating). By hypothesis, only the latter type of prediction would incur additional processing costs if the prediction is disconfirmed. Since a pre-updated prediction is integrated into the sentence representation, if it is then disconfirmed, inhibition or suppression is required in order to "override" the integrated representation and allow integration of the actual input instead (Kuperberg, Brothers, & Wlotko, 2020; Ness & Meltzer-Asscher, 2018b).

This pattern was indeed found in event-related potentials (ERP) studies which compared the responses to unpredictable words in sentences with different constraint. Predictability is commonly measured in a cloze task: participants are given truncated sentences and are asked to provide the first completion that comes to mind for each sentence; the cloze probability of a word is defined as the proportion of participants who provide it as a completion, and it is considered to reflect how predictable the word is. The constraint of a sentence is defined as the cloze probability of the most probable completion for a sentence, and it reflects the degree to which the sentence leads to a strong prediction. Several studies found that N400 amplitudes do not differ between low-cloze words in high vs. low constraint sentences. This finding suggests that the N400 reflects activation levels, associated with pre-activation, rather than a penalty for disconfirmation of strong predictions. However, low-cloze words that follow high as opposed to low constraint contexts were found to also elicit a late anterior positivity, the 'frontal post-N400 positivity' (fPNP) component (e.g. Brothers, Swaab, Traxler, 2015; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Kuperberg, Brothers, & Wlotko, 2020; Ness & Meltzer-Asscher, 2018b; see Van Petten & Luka, 2012 for a review). Thus, the disconfirmation of a strong prediction incurs additional processing costs that are not observed if no strong prediction was formed in the first place. As explained above, the pre-updating mechanism provides a plausible reason for these additional processing costs associated with the disconfirmation of strong predictions, as pre-updating entails integration into the sentence representation, which needs to be overridden to accommodate the actual input.

The association between pre-updating and commitment to a certain prediction naturally stems from the limitations of working memory (WM). Unlike long-term memory, which can simultaneously hold a seemingly infinite number of representations, much less information can be held and processed effectively in WM (e.g. "the magical number seven" suggested by Miller, 1956, "the magic number four", Cowan, 2010; Green, 2017, or even fewer items, as suggested by McElree, 2001). Therefore, it is unlikely that many competing predictions can simultaneously be pre-updated. For this reason, only a prediction that is highly activated would be integrated into the sentence's representation in WM.

1.1) Pre-updating and the P600 ERP component

The P600 component, a late positive deflection in the EEG which is maximal over posterior sites, was initially observed in response to syntactic anomalies (e.g. violation of subcategorization constraints, Osterhout & Holcomb, 1992, or agreement errors, Hagoort, Brown & Groothusen, 1993), as well as in "garden path" sentences, i.e. sentences in which syntactic reanalysis is needed (e.g. Hagoort, Brown & Osterhout, 1999; Osterhout & Holcomb, 1992; Osterhout, Holcomb, & Swinney, 1994). However, increased P600 amplitude was later observed also in grammatical sentences that do not involve reanalysis. For example, the component has been found in "semantic illusion" contexts, i.e. grammatical sentences that are semantically anomalous due to thematic role reversal or thematic violations (the "Semantic P600", see e.g. Chow & Phillips, 2013; Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Kuperberg et al., 2007).

Interestingly, increased P600 amplitude is also observed when a filler-gap dependency is completed (e.g. Felser, Clahsen, & Munte, 2003; Fiebach, Schlesewsky & Friederici, 2002; Gouvea, Phillips, Kazanina, & Poeppel, 2010; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005). For example, the verb "imitated" in a sentence such as (1a) elicits an increased P600 amplitude relative to the same verb in a sentence such as (1b) (Kaan et al., 2000). The difference between these sentences is that upon encountering the verb "imitated" in (1a), the filler ("who") is also integrated, as the complement of that verb, receiving the thematic role of theme. In contrast, the processing of the verb "imitated" in (1b) does not include this additional process of integrating a complement. Thus, there are more integration demands at the verb in (1a) relative to (1b), and these are presumably reflected in the larger P600.

(1) a. Emily wondered who the performer in the concert had **imitated** __ for the audience's amusement.
b. Emily wondered whether the performer in the concert had **imitated** a pop star for the audience's amusement.

These and other findings have led numerous researchers to argue that the P600 reflects integration processes (Brouwer, Fitz, & Hoeks, 2012; Delogu, Brouwer, & Crocker, 2019; Kaan, Harris, Gibson, & Holcomb, 2000; For further discussion regarding the functional nature of the P600, also see Chow & Phillips, 2013).

A recent study used the P600 ERP component to look for electrophysiological indication for pre-updating (Ness & Meltzer-Asscher, 2018a). In this study, participants read strongly and weakly constraining sentences in Hebrew (e.g. 2a and 2b below, respectively) while ERPs of the predictable noun ('the book'), as well as its preceding verb ('find'), were recorded.

(2) a. biglal še-ofir lo makir et ha-sifria, ha-safranit azra lo **limco** et
since that-ofir not know ACC the-library, the-librarian helped him **to-find** ACC
ha-sefer še-hu haia carix
the-book that-he COP needed
'Since Ofir isn't familiar with the library, the librarian helped him **find the book** he needed.'
b. ofir xipes ve-xipes bemešex šaot, aval lo ecliax **limco** et
ofir searched and-searched for hours, but not succeeded **to-find** ACC
ha-sefer še-hu haia carix
the-book that-he COP needed
'Ofir had searched for hours, but he couldn't **find the book** he needed.'

The crucial finding of the study was that in the highly constraining sentences (e.g. 2a), relative to the weakly constraining ones (2b), an increased P600 was observed at the verb, where the prediction was generated. This finding indicates integration of the strongly predicted word into the sentence representation prior to its appearance in the input, namely, pre-updating. The P600 in this case reflects an integration process very similar to the one discussed above with regard to the completion of filler-gap dependencies. An increased P600 is observed at the verb in (1a) relative to (1b) due to integration at this point not only of the verb, but also its complement (i.e. integrating the filler, assigning it a thematic role at the gap position). Similarly, an increased P600 amplitude is observed at the verb in (2a) relative to (2b) due to the integration of not only the verb, but also its (predicted) complement. Of course, these are different types of predictions (i.e., a grammatical prediction for a gap versus a lexical prediction for an upcoming word), but the underlying integration processes indexed by the P600 appear to be similar (e.g. thematic role assignment, phrase structure building, etc.). We note that the interpretation of the P600 effect observed at the verb in (2a) relative to (2b) does not hinge on a highly specific characterization of the processes reflected by the P600; it is compatible with many of the current accounts for the P600 component. Accumulating evidence suggests that, unlike the N400 component, the P600 is unlikely to reflect activation levels or retrieval difficulty; instead, it was argued to reflect integration processes (e.g. Brouwer, Fitz, & Hoeks, 2012; Delogu, Brouwer, & Crocker, 2019; Kaan, Harris, Gibson, & Holcomb, 2000). Thus, although the specific functional nature of the P600 component is still under debate, if any of the processes involved in integrating a word - syntactic structure building, dependency formation, thematic role assignment, semantic integration, etc. - affect P600 amplitude, this is sufficient for the conclusion that pre-updating (i.e. integrating a predicted word) is reflected in the P600 effect observed in Ness & Meltzer-Asscher (2018a), as well as in the current study.

Interestingly, the P600 pre-updating effect observed in Ness & Meltzer-Asscher (2018a) was greater for participants with higher WM performance, as measured by a reading span task. This finding suggests that the tendency to engage in pre-updating depends on this individual trait, which is consistent with the claim that pre-updating involves WM representations.

In addition, this P600 pre-updating effect on the verb was followed by greater decrease in P600 amplitude at the noun in highly constraining versus weakly constraining sentences, for participants with higher WM performance. This finding reflects the decreased integration demands when encountering a predicted word that had already been pre-updated.

1.2) When and how is pre-updating initiated? A threshold mechanism

The results discussed above provide support for the pre-updating mechanism; however, it is still not clear when and how this process is initiated. Understanding what triggers pre-updating is vital in order to establish an accurate model of how the two prediction mechanisms, namely pre-activation and pre-updating, are incorporated within the general processing stages of a word in a sentence. The current study is aimed to test the view presented in Ness & Meltzer-Asscher (2018a), shown in Figure 1.

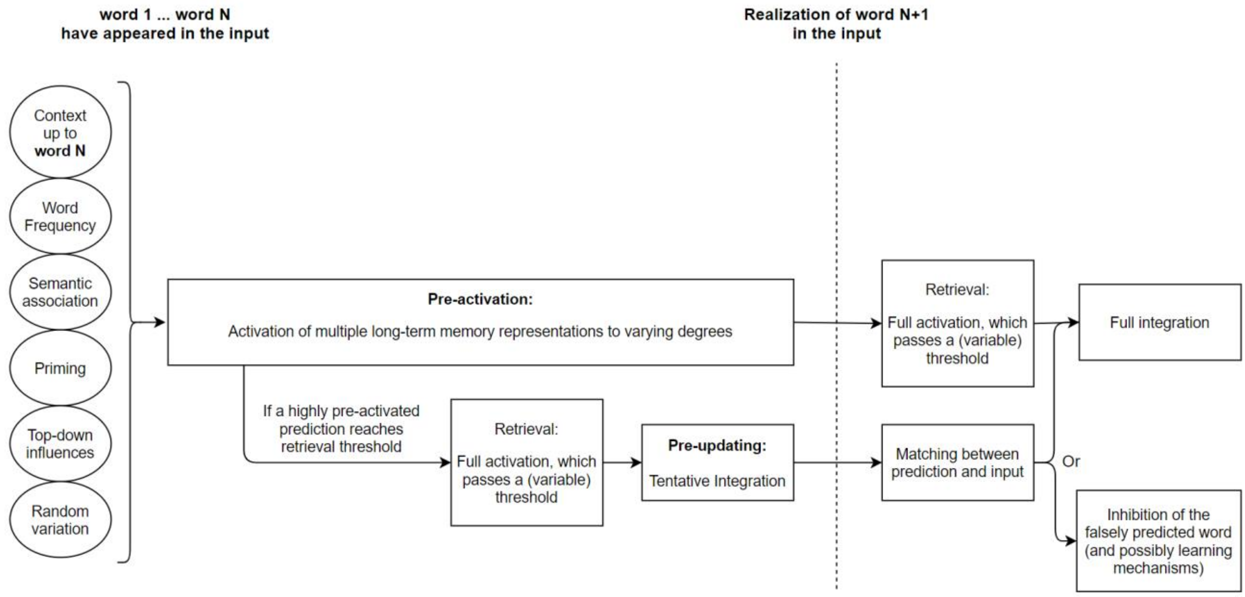

Figure 1. An outline of prediction processes within the processing stages of a word. Adapted from Ness & Meltzer-Asscher (2018a) with permission.

At every stage during sentence processing, multiple representations in long-term memory are pre-activated. Many different factors contribute to the activation level of a word: the context, lexical properties of the word (e.g. frequency), idiosyncratic influences and random noise. Once the activation level of a certain word reaches a retrieval threshold, this word is regarded as retrieved, which initiates its integration into the sentence's representation being built in WM.

In most cases, after processing word N in the sentence, no additional word passes the retrieval threshold prior to the realization of the next word, N+1. Then, bottom-up activation of word N+1 from the input will push it past the retrieval threshold, and its integration into the sentence's representation will occur. However, if after processing word N, the pre-activation of a certain word is strong enough, it can pass the threshold prior to bottom-up evidence, namely prior to the realization of the next word in the input. This is when pre-updating occurs. The strongly predicted word will be tentatively integrated into the sentence's representation, prior to its realization in the input. When the next word appears in the input, the pre-updated word will be matched against it. At this point, if the input matches the pre-updated word, integration is finalized. This stage would be less demanding than the integration of a word that had not been pre-updated. If the input does not match the pre-updated word (which would not occur very often, since pre-updating of a weak prediction would not have happened in the first place), then inhibition of the falsely predicted word is needed in order to enable integration of the actual input, incurring processing costs.

An additional aspect of the suggested mechanism is that the retrieval threshold is variable, and may shift in order to keep the balance between the benefits of confirmed pre-updates and the costs of disconfirmed ones. First, the threshold can differ between individuals, e.g. due to individual differences in WM capabilities. This can account for the higher tendency to pre-update observed for participants with higher WM performance in Ness and Meltzer-Asscher (2018a), discussed above. Secondly, the threshold may also be adjusted by top-down control, adapting to the degree to which reliance on strong prediction is beneficial in a given situation (e.g. due to task demands, noisy input, or predictive validity).

Notably, based on this view, pre-updating is not a designated prediction mechanism per-se. Any word in a sentence needs to be activated, retrieved and integrated. In order to assume that pre-updating exists, we merely need to assume that integration can occur without the need to wait for bottom-up activation, namely when top-down activation is strong enough to reach retrieval threshold. Therefore, a fundamental implication of such a mechanism is that pre-updating should occur whenever a word is activated strongly (and quickly) enough to pass the threshold prior to realization of the next word in the input. This prediction of the model is tested in the current study, by combining ERP measurements with a speeded cloze task.

1.3) The generation of cloze responses

The process of cloze response generation was recently addressed in a study by Staub and colleagues (2015). In this study, the authors conducted a timed version of the cloze task, in which participants read sentence beginnings that were presented word-by-word at a fixed rate, followed by a blank line prompting them to complete the sentence out loud as quickly as possible. The authors tested whether production onset is influenced by cloze probability, as well as by sentence constraint. The results showed that production onsets were correlated with the cloze probability of the produced response, i.e. higher cloze (more predictable) words were produced faster. More interestingly, there was also a correlation between production onset and sentence constraint. Words with similar (low) cloze probabilities were produced faster in sentences with high constraint, namely when the sentence had a strong, very probable, alternative completion, than in less constraining sentences. To illustrate, consider sentences (3)-(4) (cloze probabilities appear in parentheses)[12]. Sentence (3) is a high constraint sentence, since it has a very probable ending, "popcorn". In contrast, sentence (4) is a low constraint sentence, since it does not generate any highly probable completion. Staub et al. (2015) showed that a low cloze completion will be produced faster when the sentence is more constraining, e.g. "candy" in (3) will be produced faster than "notebook" in (4), all else being equal.

(3) Before the movie even started, the kids started to eat the…
   a. popcorn (75%)
   b. candy (10%)

(4) In the classroom, Amy opened the cabinet to take out the…
   a. book (25%)
   b. notebook (10%)

These results can be explained if cloze probability distributions do not only reflect the behavior of a population, but are also in some way reflected within each individual's mind. Even those individuals who ultimately produce a completion other than the most probable one need to have some form of latent representation of the population cloze probability distribution, in order for constraint, namely the strength of the most probable completion, to influence the production onset of other completions.

As argued by Staub et al. (2015), these results rule out two possible accounts for the generation of different cloze responses to the same sentence (see also Van Petten & Luka, 2012). First, the variability

[12] This example was constructed for explanatory purposes, and the presented cloze probabilities are rough estimates.

in cloze responses is unlikely to stem (solely) from variability in linguistic experience and real-world knowledge across the population (i.e. different individuals, having accumulated different statistics through their linguistic input, have different 'most probable' completions). Additionally, cloze responses cannot be produced by each participant simply sampling once from the cloze probability distribution, as suggested by Smith and Levy (2011)[13]. Such accounts would not predict an influence of the response's probability on its production onset, let alone an influence of the probabilities of alternative responses.

Instead, Staub and colleagues (2015) demonstrate that an activation 'race' model naturally captures the observed results. In this model, multiple possible completions are activated and 'race' towards a retrieval threshold (a specific level of activation). The cloze response ultimately produced is the one that had accumulated sufficient activation and reached this threshold first. Importantly, each response has a mean activation rate, i.e. a stable representation of the rate at which this word is activated in the context. This rate is presumably driven by the connection strength between the word and the context (i.e. how strongly the context promotes the retrieval of that word). The word with the fastest mean activation rate would most often win the race, and would thus be the word with the highest cloze probability. Nonetheless, the activation rate of each response is somewhat variable, due to random noise and idiosyncratic influences. Due to this probabilistic nature of activations, for a given participant in a given trial, a response with slower mean activation rate (i.e. a word that is less strongly connected to the context) can reach retrieval faster than the response with fastest mean activation rate. However, for this to happen in a constraining sentence, when one completion has a very fast mean activation rate (i.e. it is a highly probable word, which is strongly connected to the context), the less probable word (with slower mean activation rate) must be activated fast enough, in that trial, in order to pass the retrieval threshold prior to the more probable word. The model can thus explain why the production onsets observed for cloze responses with similar cloze probabilities are shorter in high constraint contexts than in low constraint ones. Importantly, this implies that an individual who, in a given trial, produces a low cloze response to a high constraint sentence, does not have an entirely different distribution of mean activation rates for the possible responses (i.e. they do not have a different distribution of connection strengths), compared to individuals who produce the high cloze response. Instead, the production of the low cloze response is caused by moment-to-moment variation (i.e. noise) which occasionally lead to exceptionally fast activation rate for that low cloze response, allowing it to win the race over the more probable word. We note here that we assume that the mean activation rate of a word is driven by the connection strength between that word and the context, an assumption not explicitly articulated in Staub et al. (2015). Admittedly, connection strengths may not be the only neurally plausible representation for the mean activation rates of different words by a given context. However, for clarity of presentation, we use here 'connection strengths' to refer to the stable representation of mean activation rates (which drive cloze probability distributions).

1.4) The current study

The current study is aimed to test the hypothesis that pre-updating is initiated by a noisy activation race towards a threshold (see section 1.2), similar to the generation of cloze responses. On this view, pre-updating should occur whenever a predicted word passes retrieval threshold prior to realization in the

[13] This suggestion may not be intended as a description of the actual process taking place in the brain, but rather as a computational description of the task (as also noted by Staub et al., 2015).

input. As discussed above, the results of Staub and colleagues (2015), indicate that when a speaker produces a low cloze word as the first completion that comes to mind for a high constraint sentence, this word has to have been activated fast despite its low cloze probability. Therefore, if pre-updating is indeed initiated by an activation threshold, then when a participant produces a low cloze word as a cloze response to a high constraint sentence, this word should have also been pre-updated, due to its strong activation at that moment and despite its low cloze probability.

Crucially, the results of Staub et al. (2015) also indicate that the distributions of connection strengths (i.e. mean activation rates for the possible completions for a sentence), which underlie cloze probability distributions, are in some way represented within each individual's mind (otherwise the cloze probability of an unproduced response cannot influence the production onset of another response). Despite this, a low cloze word is sometimes produced for a high constraint sentence, showing that activations sometimes diverge from these connection strengths. Thus, activations at a given moment provide a less reliable measure of the actual likelihood of a word to be the continuation of a sentence, compared to connection strengths. Notably, in the speeded cloze task, the participant's goal is to produce a completion as quickly as possible, and there is no 'correct' response. In such a task it is therefore reasonable that (possibly noisy) activations should be "good enough" to rely on, and a low cloze word will be produced as a cloze response to a high constraint sentence if it is more activated at that moment, making it easier to retrieve. This raises the question of whether the underlying information regarding the actual likelihood of a word to appear given a context (represented in connection strength) can be accessed in other tasks and by other processes, and whether this information (rather than noisy activations) is relied on in processes for which an accurate estimation of a word's likelihood to appear is more crucial.

As detailed above, pre-updating is a strong form of prediction, hypothesized to incur processing costs if disconfirmed. Therefore, the pre-updating mechanism should be triggered only by a prediction that has a high likelihood of success. The decision of whether to pre-update is thus fundamentally different from the decision of which word to produce in the cloze task: while there is no penalty for producing a low cloze completion in the cloze task, pre-updating an improbable word would often result in processing costs, since this prediction is highly likely to be disconfirmed. Since connection strength is a reliable measure of the likelihood of the prediction to be correct, rationally, pre-updating could be triggered based directly on the connection strength between the predicted word and the context, if this information is accessible, rather than based on noisy activations (even if cloze response generation is based on the latter). If this was the case, then when the activation of a low cloze word is stronger than the activation of a high cloze alternative, the low cloze word would be produced as a cloze response, without having been pre-updated prior to production. This is so since the participant's strongest prediction at that moment in time (the most activated word) was an improbable word (with relatively weak connection strength to the context) and it therefore did not trigger pre-updating. On this view, then, cloze response generation and pre-updating are based on different information types, due to their different nature. Our hypothesis, namely that pre-updating, like cloze response generation, is initiated by a race towards an activation threshold, states that this is not the case. Namely, it suggests that although our brain does have, in some form, a representation of the distribution of connection strengths (which is a more reliable measure of a word's likelihood to appear), this information cannot be (or is not) accessed directly, without the mitigation of noisy activations, even in a decision that might incur failure costs, i.e. pre-updating.

The current experiment was therefore designed to examine the prediction processes that take place prior to when a participant provides a low cloze word (e.g. 'candy') as a continuation for a high constraint sentence (e.g. 'Before the movie even started, the kids started to eat the…'). Specifically, we hypothesized that the P600 pre-updating effect would occur in high- relative to low-constraint sentences, not only when the participant predicts the high cloze word, but also when a low cloze word is predicted in the high constraint context. This would indicate, as explained above, that pre-updating is initiated by an activation race toward a threshold, similar to the generation of cloze responses.

To test this, the current experiment employed a speeded cloze task (Staub et al., 2015), but crucially accompanied by ERP measurement enabling us to target the prediction processes that precede production. Participants read sentence beginnings, ranging in constraint, and had to provide a completion out loud as quickly as possible when prompted by the appearance of a blank line. ERPs were measured on the verb where the prediction could be generated, prior to the production prompt. An additional word, the Hebrew accusative case marker 'et', was presented between the verb and the production prompt, to ensure that production does not contaminate the ERP (see Materials section). This design allowed us to analyze the electrophysiological data based on the participant's cloze response, reflecting their strongest prediction at that moment in time.

If pre-updating is indeed initiated by an activation threshold, the prediction processes when a low cloze word is produced in a high constraint sentences should be on a par with those that take place when a high cloze word is produced in these contexts, i.e. this is an instance of strong prediction, despite the low probability of the predicted word. Therefore, pre-updating should occur, resulting in an increased P600 where the prediction is generated, similar to that obtained for high cloze completions of high constraint sentences.

The current design also allowed us to replicate the behavioral results observed by Staub et al. (2015), looking at production onsets of cloze responses. We additionally administered a reading span test to assess participants' WM performance in order to replicate the correlation between the P600 pre-updating effect and WM performance, observed in Ness & Meltzer-Asscher (2018b).

2) Methods

2.1) Participants

Participants were 48 Tel-Aviv University students (14 males), all native Hebrew speakers, with an average age of 24.85 (range: 18-39). Participants were given course credit or were paid 80 NIS (~22$) for their participation. The experiment was approved by the Ethics Committee in Tel Aviv University.

2.2) Materials

The materials consisted of 156 Hebrew sentences, varying in constraint. The sentences were composed in pairs such that each pair included a high constraint sentence and a low constraint sentence (based on a cloze probability questionnaire, as detailed below). See Table 1 for example sentences. The critical word for the ERP analysis was the verb (marked in bold in Table 1), which was identical for the two sentences in each pair. The verb was always followed by the Hebrew accusative case marker ('*et*', an orthographic word), which was then followed by a blank line prompting participants to produce a completion. The Hebrew accusative case marker therefore served to distance the verb from the production prompt, to ensure that production does not contaminate the ERP. The full list of materials is provided at: https://osf.io/vjwds/?view_only=6758ee1a8b3f4e2b993b48a36cbb8afc.

Table 1: Example sentences

Constraint	Sentence frame
High	*biglal še-ofir lo makir et ha-sifria, ha-safranit azra lo* ***limco*** *et* __ since that-Ofir not know ACC the-library, the-librarian helped him **to-find** ACC __ 'Since Ofir isn't familiar with the library, the librarian helped him **find** __'
Low	*ofir xipes ve-xipes bemešex šaot, aval lo ecliax* ***limco*** *et* __ ofir searched and-searched for hours, but not succeeded **to-find** ACC __ 'Ofir had searched for hours, but he couldn't **find** __'

In order to prevent participants from anticipating when they will be prompted to produce a cloze response, the sentence fragments varied substantially in length, with the number of words prior to the critical verb ranging from two to fifteen. In addition, some of the sentences included additional verbs and accusative case markers after which the sentence continued, and the production prompt was not yet presented. The number of words prior to the verb did not differ between the high and low constraint sentences (High: M = 7.2, SD = 2.4; Low: M = 7.2, SD = 2.5; p = .794), nor did the length or frequency of the word prior to the verb (length – High: M = 4.3, SD = 1.9; Low: M = 4.4., SD = 1.4; p = .874; frequency – High: M = 1027.5, SD = 2565.6; Low: M = 1277.3, SD = 3500.8; p = .612). Presentation order was randomized for each participant. Sentences from the same pair (which contained the same verb) were separated by at least 50 trials, with the order of presentation counterbalanced between conditions.

Cloze probability questionnaires: cloze probability questionnaires were conducted on two versions of each sentence – truncated prior to the verb, and truncated following the verb. This was done to assess the cloze probability of the verb, which is the critical word for the ERP analysis, and to assess the constraint of each sentence (after the verb). Each sentence fragment was presented to at least 30 participants (different from those participating in the ERP experiment). Participants were instructed to complete each sentence with the first completion that comes to mind. Several questionnaires were conducted, but no two versions of the same sentence nor two sentences from the same pair were presented to the same participant. The order of the sentences was randomized for each participant.

Based on the results of these questionnaires, the experimental materials were constructed such that the cloze probability of the critical verb was very low in both high and low constraint sentences, and did not differ significantly between the conditions (High: M = 5.8%, SD = 15.7; Low: M = 3.2%, SD = 8.7; p = .093). The average constraint following the verb (i.e. the constraint calculated based on cloze responses generated when the presented sentence fragment included the verb) was 78.9% in the high constraint sentences (range: 50%-100%), and 26.7% in the low constraint sentences (range: 0-50%).

2.3) Procedure

Stimuli were presented using the E-prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Each trial was preceded by a 1000ms fixation cross. Sentences were presented word-by-word in the middle of the screen for 250ms, with a 350ms ISI. Following the sentence fragment, a blank line prompted participants to produce a completion. Participants were instructed to produce a completion out loud as quickly as possible once the blank line appears. The blank line was presented until the participant finished producing a completion and pressed a button to continue, or for up to 5 seconds (i.e. if the participant did not produce a completion within 5 seconds, the trial was terminated). After each trial, a string of number signs (####) appeared on the screen and the participant pressed a button when they were ready to start the next trial. Participants were encouraged to take as many breaks as needed. Prior to the experiment, participants completed a practice block of five trials. The experimental session (including EEG set-up) took 60-90 minutes.

Reading span test: To assess WM performance, each participant completed a reading span test. The test was performed after the main experiment. The test's procedure was based on Daneman and Carpenter (1980), with minor differences. Participants read aloud series of Hebrew sentences, after which they had to recall the last word of each sentence. The number of sentences in the series increased from two to six. Participants had three series in each level, and the last level at which a participant correctly recalled all words in at least two series was defined as this participant's reading span (i.e., when the participant failed to recall a word in two series of the same level, the test was terminated and the participant's reading span was set at the preceding level). Two practice series (at the two-sentence level) were performed prior to the test, in which participants could make mistakes and ask questions.

2.4) EEG recording and pre-processing

The electroencephalogram (EEG) data were recorded using a BrainVision actiCap system with 32 Ag/AgCl scalp electrodes attached according to the 10-20 system. Two electrodes were used to monitor EOG, located at the infraorbital ridge and the outer canthus of the right and left eyes respectively. Electrode impedances were kept below 5 kΩ for all scalp electrodes and below 15Ω for the EOG electrodes. During recording, the EEG was referenced to Fp2. The EEG was then re-referenced offline to the average of the left and right mastoid electrodes. Data were collected at a 250 Hz sampling rate and low-pass filtered at 70 Hz. Data were then bandpass-filtered between 0.1 and 30 Hz, and segmented into 1200 ms epochs, including -200 to 1000 ms relative to the onset of the critical word. The 200 ms prior to the onset of the critical word were used for baseline correction. Trials contaminated by blinks, eye movements, excessive muscle activity or amplifier blocking were rejected off-line before averaging and excluded from further analysis (this affected 4.78% of the trials).

2.5) EEG data analysis

Based on the typical time-window of the P600 component, mean amplitudes over the 500-800ms time-window were analyzed. Electrodes were grouped based on their anteriority and laterality (Anterior - Left: F7, F3, Fp1, FC5, FC7, T1, C3; Middle: Fz, Cz; Right: F8, F4, Fp2, FC2, FC6, T7, C4; Posterior - Left: P7, P3, O1, CP5, CP1; Middle: Pz, Oz; Right: P8, P4, O2, CP2, CP6) in order to reduce the number of comparisons and the familywise error rate (see Luck, 2014) while still allowing to assess the topography of the effects. High constraint trials were divided based on the produced word, forming three conditions: HH (High constraint, High cloze word produced), HL (High constraint, Low cloze word produced), L (Low constraint). This resulted in a repeated-measures ANOVA with the factors Anteriority (Anterior, Posterior), Laterality (Left, Middle, Right), and Condition (HH, HL, L). The Huyhn-Feldt adjustment for nonsphericity of variance was applied when the sphericity assumption was violated. In these cases, the corrected p-value is reported with the original degrees of freedom.

2.6) Audio recordings – transcription, onset measurement and data analysis

Production onset was marked using DeepWDM, a recurrent neural network for word duration measurement (Goldrick et al., 2018). A coder then listened to each of the recordings, transcribed the production, and corrected the marked onset when needed (17.75% of the trials). Trials with speech errors, repairs or filled pauses were excluded (1.7% exclusion). Singular/plural and masculine/feminine forms of the same noun were collapsed by coding each noun as its singular-masculine form (e.g. 'shirts' and 'shirt' where counted as the same response). When the produced response consisted of more than one word, only the first word was coded unless the words formed a compound. Since the presented sentence fragments always ended with the Hebrew accusative case marker 'et', which only precedes definite nouns, the vast majority of produced responses began with a definite determiner ('the' - *ha)*. Responses that included other determiners (e.g. *kol ha-kelim* – 'all the-tools') were coded without the determiner (such responses were very rare).

Production onsets were analyzed with linear mixed-effects models. Constraint and cloze probability were treated as continuous variables. Analyses were conducted using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2014) in the R software environment (R Development Core Team, 2011). Data were winsorized by replacing data points exceeding 2.5 standard deviations (SD) from each participant's mean with the value of 2.5 SDs from that participant's mean (affecting 4.0% of the data). All models initially included the maximal random effects structure for subjects (i.e. intercept and slopes of all fixed effects and interactions). The random effects structure was reduced when necessary in order to achieve convergence (the reduced models are specified in the Results section), by iteratively removing the random slope associated with the smallest variance (Barr et al., 2013). Random effects for items were not included since they could not be estimated independently of the fixed effects, as each item occurred at only one level of Constraint (as was done in Staub et al., 2015).

3) Results

The data are provided at: https://osf.io/vjwds/?view_only=6758ee1a8b3f4e2b993b48a36cbb8afc.

3.1) Behavioral results

To analyze production latencies, we followed the analysis reported in Staub et al. (2015). The full results of the analyses are reported in Table 2. We first fitted a model that included cloze probability as the only fixed effect. Higher cloze probabilities led to shortened production onsets ($p < .001$). We then fitted a second model, which included constraint as a fixed factor, as well as a factor indicating whether or not the produced word was the modal response, namely the most probable response for that item. A significant effect of constraint was found ($p < .001$), such that higher constraint led to shortened production onsets. Additionally, modal responses were produced significantly faster than non-modal responses ($p < .001$), indicating that the effect of cloze probability in Model 1 cannot be solely driven by constraint, as constraint was accounted for in Model 2 (see Figure 2a). We then fitted a third model, which included cloze probability and constraint as fixed factors, and only included responses with cloze probability <.4 (which constituted 55.9% of the data), allowing to separate constraint and cloze probability.[14] In order to achieve convergence, all random slopes had to be removed from this model.

[14] For responses with cloze probability of .5 or higher, cloze probability and constraint are perfectly correlated (the constraint is always the cloze probability of the produced word in these trials). Similar to Staub et al. (2015), we had very few responses with cloze probability between .4 and .5 that were not the modal response (when

Significant effects of both cloze probability and constraint were found ($p < .001$ for both effects), confirming that the production of low cloze completions is indeed influenced by sentence constraint in addition to cloze (See Figure 2b).

The behavioral results thus replicated the findings of Staub and colleagues (2015), indicating that words with higher cloze probabilities are produced faster, and that words with similar cloze probabilities are produced faster when the sentence constraint is higher.

Table 2: Mixed-effects regression models coefficients for production onsets

	Estimate	SE	df	t value	p value
Model 1					
Cloze probability	-0.236	0.016	46.86	-14.35	< .001
Model 2					
Constraint	-0.176	0.015	46.43	-11.89	< .001
Modal	-0.201	0.021	43.52	-9.77	< .001
Constraint : Modal	0.049	0.021	48.75	2.36	.022
Model 3					
Cloze probability	-0.170	0.011	7332	-15.80	< .001
Constraint	-0.136	0.012	7330	-11.38	< .001
Cloze probability : Constraint	0.004	0.015	7326	0.33	.741

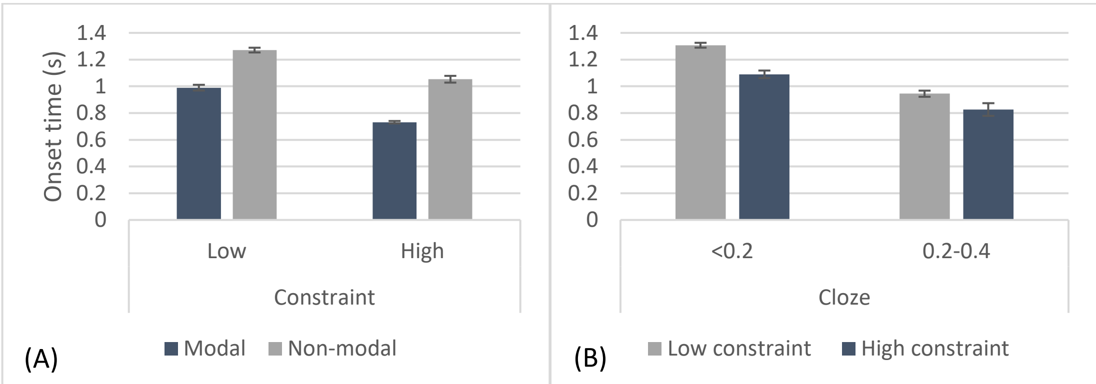

Figure 2. (A) Mean onset time for modal and non-modal responses, in sentences with high (>0.5) and low (<=0.5) constraint. (B) Mean onset time by cloze probability, in sentences with high (>0.5) and low (<=0.5) constraint. Error bars represent standard error of the mean.

3.2) EEG results

High constraint trials were divided based on the produced word, forming three conditions: HH (High constraint, High cloze word produced; this condition included 2759 trials overall, after exclusions), HL (High constraint, Low cloze word produced, 947 trials), L (Low constraint, 3670 trials). Mean amplitudes over the 500-800ms time-window relative to the onset of the verb were entered into a repeated-measures ANOVA with the factors Anteriority (Anterior, Posterior), Laterality (Left, Middle, Right), and Condition (HH, HL, L). Grand averaged ERPs and scalp distributions of the effects are displayed in Figure 3. The full ANOVA results are provided in Table 3.

the modal response is produced, constraint is again equal to cloze probability), and therefore decided to include only responses with cloze probability < .4 in this model (as in Staub et al. 2015).

500-800ms time window: The results showed a significant effect of Condition ($p < .001$). Additionally, a significant interaction was found between Anteriority and Condition ($p = .001$), such that the effect of Condition was greater in posterior electrodes, consistent with the common topography of the P600 component. Pairwise comparisons indicated greater positivity in the HH condition relative to the L condition ($F(1,47) = 17.331$, $p < .001$), as well as greater positivity in the HL condition relative to L ($F(1,47) = 20.206$, $p < .001$). The HH and HL conditions did not differ significantly ($F(1,47) = 0.037$, $p =.848$). These results indicate a P600 effect in the two high constraint conditions – namely whether the produced response was the high cloze response or a low cloze one – relative to the low constraint condition.

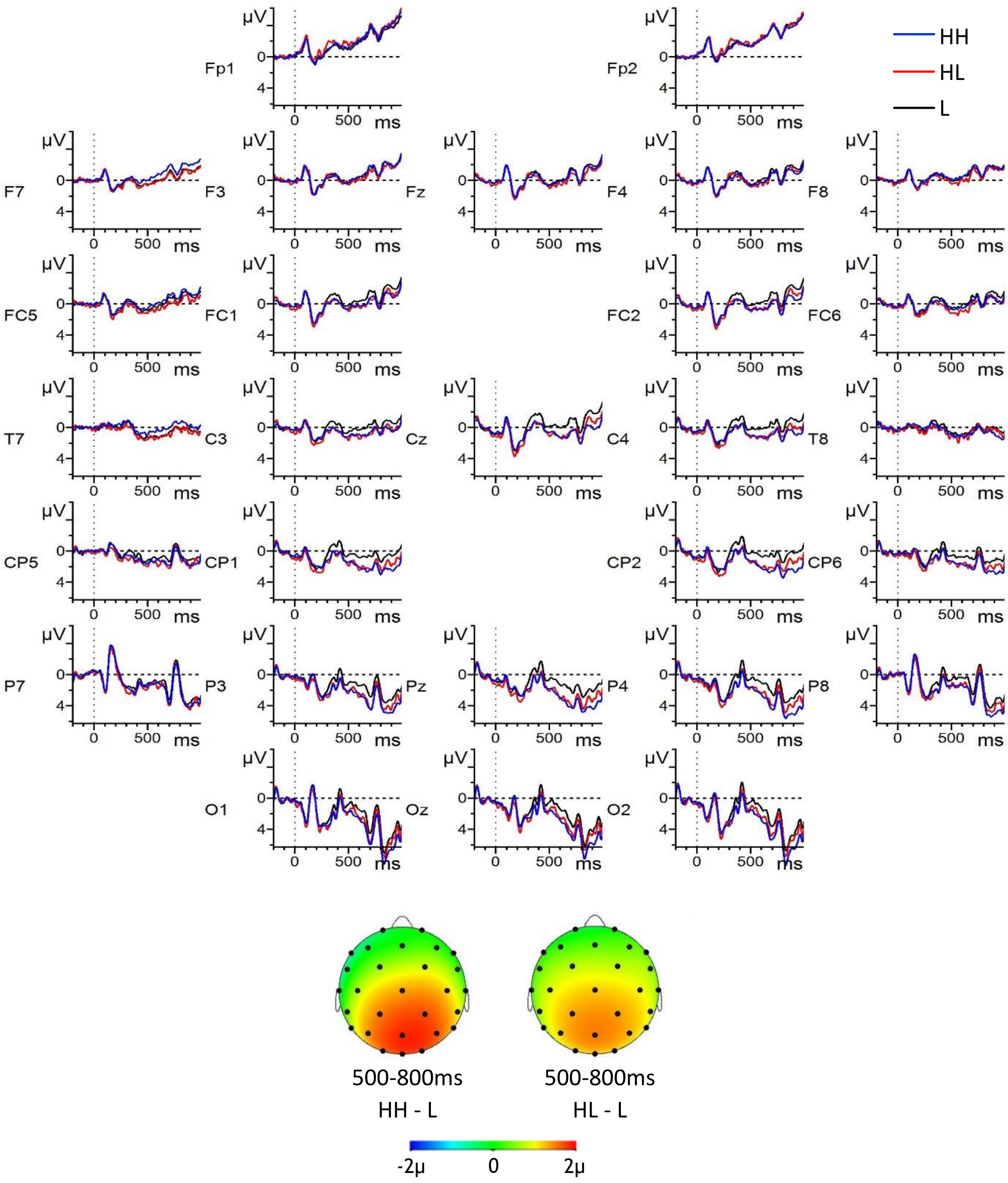

Figure 3. Grand averaged ERPs and scalp distributions of the P600 effects (500-800ms relative to the verb onset). HH = High constraint, High cloze word produced; HL = High constraint, Low cloze word produced; L = Low constraint.

Table 3: ANOVA results for ERP data

	df	F value	p value
P600			
Condition	2,94	10.63	< .001
Anteriority	1,47	96.05	< .001
Laterality	2,94	22.12	< .001
Condition : Anteriority	2,94	7.52	.001
Condition : Laterality	4,188	8.19	< .001
Anteriority : Laterality	2,94	3.58	.041
Condition : Anteriority : Laterality	4,188	0.55	.661
N400			
Condition	2,94	9.99	< .001
Anteriority	1,47	33.05	< .001
Laterality	2,94	9.46	< .001
Condition : Anteriority	2,94	4.11	.020
Condition : Laterality	4,188	3.21	.023
Anteriority : Laterality	2,94	0.33	.707
Condition : Anteriority : Laterality	4,188	0.41	.750

Additionally, a by-trial correlation was found between P600 amplitude (i.e. mean amplitude over 500-800ms relative to verb onset) and production onset, such that faster productions occurred in trials with larger P600 amplitudes at the preceding verb (Pearson correlation = -.048, N = 7022, $p < .001$. See figure 4). Finally, a by-participant correlation was found between the P600 effect (i.e. the mean difference in amplitude between the high and low constraint sentences, 500-800ms relative to verb onset) and reading span (Pearson correlation = 0.295, N = 48, $p = .042$. See Figure 5), replicating previous results (Ness & Meltzer-Asscher, 2018b).

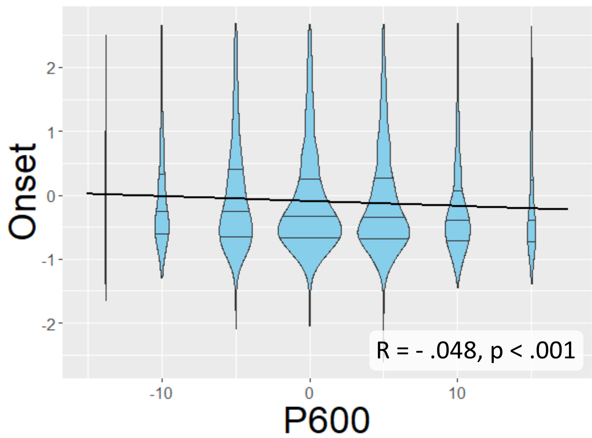

Figure 4. By-trial correlation between P600 amplitude (i.e. mean amplitude over 500-800ms relative to verb onset) and production onset.

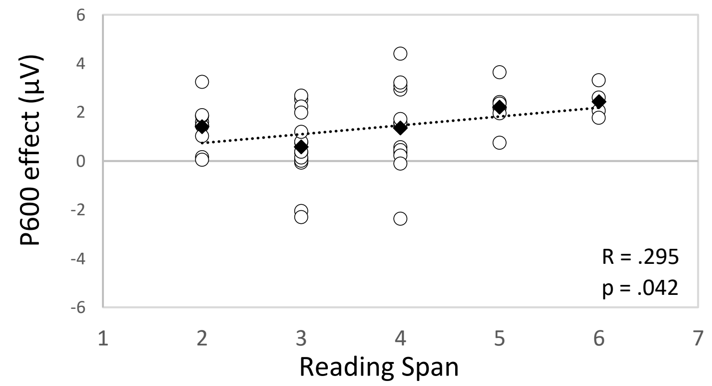

Figure 5. By-participant correlation between the P600 effect and reading span. The P600 effect was defined as the mean difference in amplitude between high and low constraint trials in the 500-800ms time window relative to verb onset. ○ – single participant, ♦ - average across reading span score.

300-500ms time window: Visual inspection of the ERPs suggested a difference between the conditions in this time window, typical of the N400 component. We therefore chose to conduct and report this analysis for completeness (see Discussion for interpretation of these results). The analysis showed a significant effect of Condition ($p < .001$). Additionally, a significant interaction was found between Anteriority and Condition ($p = .020$), such that the effect of Condition was greater in posterior electrodes, consistent with the common topography of the N400 component. Pairwise comparisons indicated greater negativity in the L condition relative to the HH condition ($F(1,47) = 15.533$, $p < .001$), as well as greater negativity in the L condition relative to the HL condition ($F(1,47) = 15.389$, $p < .001$). The HH and HL conditions did not differ significantly ($F(1,47) = 0.121$, $p = .730$). These results indicate an N400 effect in the low constraint condition relative to the two high constraint conditions.

3.3) Reliability of cloze data

Since we had cloze responses on the same 156 sentences from two datasets – the main experiment (a speeded cloze task) and the pre-test (an offline cloze questionnaire), we also carried out an analysis to assess the reliability of cloze data and the influence of the procedure (i.e. speeded vs. offline) on the data. Overall, responses were extremely similar in both datasets. The modal response to the high constraint sentences differed between the two datasets only in one item, and in this item the constraint was relatively close to 50% and the words that alternated were highly related near-synonyms (*ha-metupal* – 'the patient', and *ha-xole* – 'the sick person'). The overall average constraint was also very close in the two datasets (52.83% in the pre-test, 52.72% in the main experiment) and the by-item correlation between sentence constraint calculated in each dataset was very high ($N = 156$; $r = 0.90$, $p < .001$).

4) Discussion

The current study investigated what triggers the pre-updating mechanism. We hypothesized that pre-updating is triggered by an activation race toward a retrieval threshold. This means that similarly to the decision which cloze completion to produce, the decision of whether to pre-update would be noisy rather than based directly on the strength of the connection between the predicted word and the context, leading to occasional pre-updating of improbable (low cloze probability) words, despite their high likelihood of being contradicted by the upcoming input.

Looking at the processes involved in the prediction of a low cloze word in a high constraint context, we asked whether the P600 pre-updating effect would occur in high- relative to low-constraint sentences, not only when the participant predicts the high cloze word, but also when they predict a low cloze word. In such cases the participant's strongest prediction, reflected in their cloze response, differs from the population's most predictable (most probable) word.

Participants completed sentences ranging in constraint, and ERPs were measured on the verb where the prediction can be generated, prior to when the participants were prompted to produce a completion. The behavioral results indicated unique contributions of both cloze probability and sentence constraint to the production onset of cloze responses. Higher cloze probability reduced production onset times above and beyond sentence constraint, and vice versa. These results replicate the results of Staub and colleagues (2015) in a different language and with a different set of materials.

The ERP results showed an increased P600 amplitude at the verb in high constraint sentences relative to low constraint ones. This effect replicates the results of Ness & Meltzer-Asscher (2018a). As discussed in the Introduction, this effect was suggested to reflect pre-updating, indicating increased integration costs when integration at the verb includes not only the verb itself, but also the predicted noun (in high constraint sentences). Notably, the participants' task in the current study was sentence completion. However, we believe the observed ERP effects are unlikely to be related to production. First, the P600 effect in the current study resembles in timing and topography to the P600 effect in the previous study (Ness & Meltzer-Asscher, 2018a), which did not involve production, but merely reading for comprehension. More importantly, the materials in the current study were constructed to ensure that production does not contaminate the ERPs elicited by the critical verb. This was done by having the Hebrew accusative case marker ('et') presented between the verb and the production prompt, distancing the verb from production onset, as well as by varying the sentence fragments' length (the number of words prior to the critical verb), and including additional verbs and accusative case markers after which the sentence continued and the production prompt was not yet presented (see Materials section), thus preventing participants from anticipating the presentation of the production prompt. Thus, production preparation was unlikely to have taken place at the critical verb.

The novel and crucial finding in the current study was that increased P600 amplitude at the verb in high constraint (relative to low constraint) sentences was observed even in trials in which a low cloze word was produced. This result provides evidence that pre-updating of a predicted word occurred in high constraint sentences, regardless of whether the produced cloze response (i.e. the participant's strongest prediction at that moment in time) was a high or low cloze response.

The results also replicate the correlation between the size of the P600 pre-updating effect at the verb and the participant's WM performance, such that participants with higher reading span scores showed a greater effect (Ness & Meltzer-Asscher, 2018a). This suggests that better WM capabilities allow for a greater tendency to pre-update, supporting the involvement of WM representations in this process. Additionally, the results showed a reduction in N400 amplitude in the high constraint sentences relative to the low constraint sentences. Since we did not have any hypothesis regarding the N400 component, the analysis of the N400 time window was not planned and should be regarded as exploratory. We believe that the reduction in the N400 amplitude in the high relative to the low constraint sentences is likely due to the difference in the verb's cloze probability. Although we attempted to control for verb cloze probability (see Materials section), the verb's cloze probability in the high constraint sentences was slightly higher than in the low constraint sentences (5.8% vs. 3.2%). The cloze probabilities of the verbs in all of the materials was low, which may explain why the small difference in cloze probability manifested in the ERP results: small differences in the lower range of the cloze probability scale are known to have relatively large effects on processing difficulty (e.g. Smith & Levy, 2013). Importantly, however, there is no ground to assume that it is this difference in the verb's cloze probability which caused the P600 effect. First, countless studies manipulating cloze probability have not resulted in P600 effects attributed to cloze probability (e.g. Delong, Urbach & Kutas, 2005; Kutas & Hillyard, 1984; Wlotko & Federmeier, 2012). Secondly, the direction of the P600 effect is such that the amplitude is higher in the high constraint sentences (by hypothesis due to pre-updating the upcoming noun). However, the verb's cloze probability in these sentences was higher, not lower. If anything, we would expect the integration of a more predictable verb to be facilitated. Assuming that the P600 reflects integration costs, an increased P600 due to a more predictable verb is highly unlikely, pointing to the conclusion that the P600 here indexes integration costs of the predicted, rather than the current, word. An additional affirmation that the effect in the N400 time window and the effect in the P600 time window are not driven by the same cause is that while the P600 effect was correlated with participants' reading span scores, the N400 effect was not (Pearson correlation = 0.126, N = 48, p = .392).

4.1) Pre-updating and a threshold mechanism

As explained in the Introduction, pre-updating is the integration of a predicted word into the sentence representation being built online in WM. Since WM is limited, it is unlikely that many competing predictions can simultaneously be pre-updated. Additionally, the disconfirmation of a pre-updated prediction was suggested to incur additional processing costs beyond what is needed when pre-updating did not take place, since the integrated (i.e. pre-updated) prediction has to be inhibited in order to enable integration of the unexpected word that appeared in the actual input (Ness & Meltzer-Asscher, 2018b). Rationally, pre-updating should therefore only be initiated when a prediction has a high probability to be correct.

In Ness & Meltzer-Asscher (2018a), a threshold mechanism was postulated as the trigger for pre-updating. Under this view, multiple predictions are pre-activated, but in most cases such activations are not enough for any word to be retrieved (and integrated). Only once bottom-up activation is added (i.e. the word appears in the input), the retrieval threshold is reached and the word can be integrated into the sentence's WM representation. However, if a certain word is strongly activated, for example in a highly constraining sentence, pre-activation can be sufficient for the word to reach the retrieval threshold prior to its realization in the input. This predicted word is then integrated into the sentence's WM representation, i.e. pre-updating occurs. If this is indeed the mechanism that initiates pre-updating, then pre-updating should occur whenever a word passes retrieval threshold prior to realization in the input. This means that whether or not an activated word is pre-updated should not directly depend on its cloze probability, but rather on how strongly it is activated within the participant's mind at a specific moment in time.

The results of Staub and colleagues (2015) indicate that when a participant produces a low cloze word as a completion for a high constraint sentence in the (speeded) cloze task, this word has to have been highly activated despite its low cloze probability. Therefore, our hypothesis in the current study was that if indeed an activation race towards a threshold is what triggers pre-updating, then pre-updating should occur in high constraint sentences regardless of the cloze probability of the word predicted and produced by the participant at that specific moment in time. Our results are in line with this prediction, as the P600 pre-updating effect was observed in high constraint sentences prior to both high- and low-cloze completions.

Interestingly, this is a seemingly faulty mechanism for the initiation of pre-updating. As discussed in the Introduction, the results of Staub et al. (2015) indicate that the distribution of connection strengths which underlie cloze probability distributions is in some way represented within each individual's mind (i.e. an individual who, in a given trial, produces a low cloze response to a high constraint sentence, does not have an entirely different distribution of connection strengths compared to individuals who produce the modal response). Therefore, a 'better' mechanism would take into account that pre-updating, unlike cloze response generation, is "risky", and would initiate pre-updating directly based on the connection strength of the predicted word, which provides a more reliable measure for its likelihood to appear in the sentence, rather than on noisy activations. Such a mechanism would entail that when a participant produces a low cloze word as a speeded cloze response to a high constraint sentence, i.e. when activation levels diverge from the cloze probability distribution and a low cloze word is activated more strongly than a high cloze alternative, pre-updating would not occur. Our results show that this is not the case, thus supporting a seemingly maladjusted mechanism for the initiation of pre-updating by an activation race towards a threshold, rather than by direct access to the information represented by connection strengths.

The by-trial correlation found between P600 amplitude (at the verb) and production onset of the cloze response supports our main finding by providing additional indication that pre-updating occurs when a predicted word passes retrieval threshold fast enough. We note, however, that although this correlation was statistically significant, it was very small and should therefore not be taken as a strong evidence.

4.2) The representation of cloze probably distributions

In the current study, we make the distinction between the predictability of a word given a sentence at two levels of representation: i) the strength of connection between the word and the context, which we believe reflects the participant's relatively stable estimate of the likelihood of a word to appear given the context; and ii) the strength of activation of the word at a given moment (following the context), which we believe reflects the participant's less accurate (but more accessible) approximation for the likelihood of a word to appear. However, in the psycholinguistic literature, the predictability of a word given a context is most often operationalized by the measure of cloze probability. This raises the question of the relationship between cloze probability, connection strengths and activation levels (for a similar discussion of the relation between cloze probability and predictability see Staub et al., 2015). In order to understand this relationship, we need to consider how cloze probability is measured. Cloze probability data are collected in the cloze task, by a population of participants producing cloze responses; each of these responses is the result of an activation race within a participant's mind, which is in turn driven by the participant's representation of connection strengths between the context and the potential responses (driving mean activation rates). Thus, each response represents the most activated word for a participant at a specific moment in time. Although activations are influenced by noise, on average (across participants) the stronger a word's connection to the context is, the faster its mean activation rate would be, resulting in a higher cloze probability. This may seemingly suggest that cloze probability is equivalent to connection strength. However, as pointed out by an anonymous reviewer, since cloze probability values are equivalent to "win percentages" in the race toward retrieval in the cloze task, the cloze probability of a word is not purely a reflection of the word's "strength" in the race. Instead, it is also influenced by who the other competitors are. For example, a word with a 10% cloze probability that wins over a very strong competitor 10% of the times in some sentence, is stronger than a 10% cloze probability word that wins over weak competitors 10% of the times in some other sentence. Therefore, when we measure cloze probabilities, we do not directly measure the connection strength between a word and a context (driving the word's mean activation rate). Rather, we measure a value that is correlated with this representation, but is also affected by the alternative predictions for the sentence (i.e. the other words which compete in the same race).

Importantly, this view of the mental representation of cloze probability distributions does not alter our main conclusion, that pre-updating is initiated based on noisy activations rather than the more stable and reliable representation of connection strengths. For example, let us consider a high (70%) constraint trial in which a 10% cloze probability word is produced, and a low (30%) constraint trial in which a 10% cloze probability word is produced. Admittedly, although the cloze probability of the produced word in both trials is 10%, the connection strengths in the participants' mind likely reflects stronger association between the word and the preceding context for the 10% word in the high constraint sentence relative to the 10% word in the low constraint sentence, or their win percentages would not be the same. However, this 'stronger' 10% cloze probability word (in the high constraint sentence) is still not 'strong', in absolute terms; it is somewhat stronger than the 10% cloze probability word in the low constraint sentence, but it is still much weaker than its high cloze probability competitor. Additionally, the representation of possible completions for the high constraint sentence includes the fact that there is a much stronger alternative (the 70% cloze probability word), than the produced 10% cloze probability word. If the pre-updating mechanism were initiated based on representation of connection strengths, then in the high constraint trial, despite the high activation level of the 10% cloze probability word, this word would not be pre-updated since there is indication that this prediction is improbable (much less probable than its high cloze alternative), and is unlikely to be correct. Instead, the current results indicate that when a low cloze word is highly activated (i.e. when it wins over a high cloze alternative and is produced in the cloze task as a completion for a high constraint context), it is pre-updated.

4.3) Is noisy better, good enough, or inevitable?

As explained above, the current results indicate that pre-updating is triggered by a noisy activation race rather than directly by connection strength, despite the fact that connection strength would be a better measure of the likelihood of a prediction to be correct. Pre-updating being driven by noisy activations is what gives rise to the phenomenon observed in the current study, i.e. that an improbable word can be pre-updated. The question then arises, why would pre-updating not be driven directly by the information contained in connections strength, avoiding the occasional pre-updating of an improbable prediction?

We propose three possible explanations for this puzzle. First, it is possible that due to biological properties of our brain, it may not be possible for a cognitive process to be directly informed by connection strengths. Hence, noisy activations are used simply since they are the best approximation available, given that the information contained in connection strengths cannot be accessed in a more reliable way. Secondly, even if a cognitive process can in principle directly access information contained in connection strengths, eliminating the addition of noise may require more effort (e.g. inhibiting noise sources) or more time. This would mean that noisy activations are used because they are "good-enough" – they may provide a sufficiently good approximation of cloze probability so that it would not be beneficial to exert additional efforts only to avoid the occasional cases in which a low cloze word is inadvertently pre-updated (i.e. the effort needed to access the more accurate information may outweigh the costs incurred by the disconfirmation of these pre-updates). A third, perhaps more interesting, explanation may be that noise is not an architectural flaw, but an advantage. In recent years, numerous studies have demonstrated that noise (both endogenous and exogenous) can be beneficial to neural processing. The benefits of an adequate amount of noise were shown in experiments ranging from single cell recordings to human behavior (for reviews see Guo, Perc, Liu, & Yao, 2018; McDonnell & Ward, 2011; Moss, Ward, & Sannita, 2004). For example, noise was shown to improve performance in perception, learning and decision making (e.g. Gureckis & Love, 2009; Kitajo et al., 2003; van der Groen & Wenderoth, 2016). Low endogenous neural noise was even suggested to cause the behavioral features of autism spectrum disorder (Davis & Plaisted-Grant, 2015).

Admittedly, a certain amount of noise is likely inevitable in any cognitive process. However, noise levels can be regulated, and can differ between systems (see Moss, Ward, & Sannita, 2004 for a discussion of noise regulation). This means that noise levels in a given cognitive mechanism can be optimized. Notably, in the case of the race mechanism discussed in the current study, noise levels seem quite significant. For example, let us consider a sentence that, based on accumulated language exposure, has a continuation with 70% probability of appearing, and alternative continuations with 20% and 10% probabilities. In a system with no noise at all, the word that has a 70% probability would be predicted and produced as a cloze response by any participant at any moment in time, since without noise this word would always be the most activated and would always be retrieved. As opposed to this scenario, in the cloze task we see that the modal response for such a sentence 'loses the race' 30% of the time, ergo, noise influences the final result of the race to a non-negligible degree. This may hint that noise is not merely inevitable in this system, but it has an objective. For example, if the confirmation and/or disconfirmation of pre-updated predictions triggers learning, noise can enhance the precision of the learned cloze probability distribution, strengthening the discrimination in the lower end of the distribution (i.e. the system would be better at learning that word X has a 20% probability and word Y has a 10% probability). The specific objectives of the noise in this system remains to be studied. One avenue to study this may be to look at the consequences of individual differences in the noise level in this system, which may be measured by the proportion of the trials in which a certain individual provides a low cloze response to a high constraint sentence (i.e. more noise in this system should lead a participant to provide modal responses less often).

4.4) Cloze probability and differential noise levels

Staub and colleagues (2015) suggested a computational implementation of the race model for generating cloze responses. In this model, each possible cloze response accumulates activation at a certain rate, and the finishing time for a cloze response is the time it takes for the activation of this response to reach retrieval threshold. In each cloze trial, a finishing time for each possible response is drawn from the finishing time distribution of that response, and the response with the shortest finishing time is produced. This means that the cloze probability of a certain response is the proportion of trials in which it had the shortest finishing time. The finishing time distributions of the possible responses differ in their mean. A response with a lower mean finishing time would be more likely to 'win the race' and be produced, resulting in a higher cloze probability.

The variance of a response's finishing time distribution represents the trial-to-trial variability in the response's activation rate, due to noise. In the simulations conducted by Staub et. al. (2015), a fixed standard deviation was set for all possible responses. This means that the noise level was assumed to be the same for all cloze responses, regardless of their mean finishing time or cloze probability. However, recent simulations show that this may not be the case. Nakamura and Phillips (2020) have argued that uneven noise, namely greater variability of finishing times for responses with higher mean finishing time (lower cloze probability), better captures the empirical results. This means that the finishing time distribution of low cloze responses is wider than that of high cloze responses.

One implication of this suggestion is that while the finishing times of a high cloze response would very rarely be far from its mean, a finishing time far from the mean is less rare for a low cloze response (since its finishing times distribution is much wider). Therefore, when a low cloze word is produced as a cloze response to a high constraint sentence, this is likely not because the finishing time of the high cloze alternative was much longer than its mean finishing time, allowing the low cloze word to 'win the race' despite a relatively long finishing time, but rather since the finishing time of the produced low cloze response was much shorter than its mean finishing time, overtaking the high cloze alternative.

Our results are in line with the suggestion of Nakamura and Phillips (2020). In the current study, the amplitude of the P600 pre-updating effect was similar in the two high constraint conditions. This result would only be expected if in most trials in which a low cloze probability word was produced as a cloze response to a high constraint sentence, this word won the race because it was highly activated, and was therefore also pre-updated. If instead we assume that a high cloze word is just as likely to have a finishing time far from its mean as a low cloze word is, then in a significant portion of the trials in this condition the activation of the low cloze word was not exceptionally high but it won the race because of exceptionally low activation for the high cloze alternative. This assumption would thus predict a smaller P600 pre-updating effect in this condition relative to when a high cloze word is produced, since pre-updating would only occur in a portion of the trials in the high constraint low cloze condition, compared to most or all of the trials in the high constraint high cloze condition. Importantly, the lack of difference between the P600 pre-updating effect in the two high constraint conditions is a null result, hence the current study does not provide a statistically significant result supporting differential noise levels. Our results are thus compatible with this suggestion, but further research is needed in order to provide corroborating evidence.

4.5) Previous findings on the generation of predictions

In the current study, the critical word for the ERP analysis was the verb after which a highly probable word was available (in the high constraint sentences). This means that we did not look at the processing of predictable or unpredictable words, but on the processes that take place prior – when the prediction is generated. A few previous studies have looked at the stage of generating predictions (Li, Zhang, Xia, & Swaab, 2017; Rommers et al., 2017; Fruchter et al., 2015, Ness & Meltzer-Asscher, 2018a; Ding, Wang, & Yang, 2020)[15]. As discussed above, our current results replicate the P600 pre-updating effect observed

[15] Several studies looked at the word prior to a (potentially) predictable word using a pre-nominal article that is or is not consistent with the predictable noun (e.g. Delong et al. ,2005; Martin et al., 2013; Niewland et al., 2017; Nicenboim, Vasishth, & Rösler, 2019; Szewczyk & Wodniecka, 2020; van Berkum et al., 2005; Wicha, Moreno, & Kutas, 2004). However, at the article the prediction has already been generated. These studies

in Ness and Meltzer-Asscher (2018b). However, different effects were observed in earlier studies (Li, Zhang, Xia, & Swaab, 2017; Rommers et al., 2017; Fruchter et al., 2015). Notably, the design and materials in these studies differed considerably from the current study, and from one another, leading to these inconsistent results. A discussion of these differences, and their bearing on the results, can be found in Ness & Meltzer-Asscher (2018a).

A more recent study (Ding, Wang, & Yang, 2020) is seemingly very similar to the current study. As in the current study, high and low constraint sentences were compared and ERPs were measured on a verb prior to the predictable noun. However, in contrast to the P600 pre-updating effect we observed, the results of Ding et al. (2020) showed a sustained anterior negativity (SAN). Crucially, however, the experiment by Ding et al. (2020) included an additional manipulation, the emotional content of the verbs. Therefore, the high and low constraint sentences had two versions each, one with an emotionally positive/negative verb, and one with an emotionally neutral verb. This meant that the materials had to be constructed in a way that would allow for the manipulation of the verb not to have an influence on the predictability of the following noun. i.e. the prediction of the noun in the high constraint sentences had to be based only on semantically related words prior to the verb, and the verb itself did not contribute much to the prediction. This can explain why a SAN was observed: this effect may reflect holding a prediction that is expected to be realized at a later point, resulting in increased memory load (Ness & Meltzer-Asscher, 2018a). A similar result was observed when a prediction regarding an upcoming noun had to be held in memory across an (unpredictable) adjective in Li, Zhang, Xia, and Swaab (2017).

On the other hand, in the vast majority of our materials, the prediction could not be generated before the critical verb and independently of it. For example, in a sentence such as 'The bridesmaids waited breathlessly for the bride to <u>wear</u> the **dress**', although a broad prediction related to weddings can be formed prior to the verb, at this point the sentence can continue in several ways, causing activation of several different nouns (e.g. throw the **bouquet**, walk down the **aisle**, cut the **cake**, and so on). Only once the verb appears, a specific, strong prediction (of 'dress') can be made. Additionally, this prediction can be integrated immediately, since no additional content words appear between the formation of the prediction and its expected position. Hence, we would not expect a SAN effect in our study since there is no need to hold a prediction. Instead, we expect the P600 effect reflecting the integration of the predicted noun, as indeed observed in our results.

As pointed out in Ness & Meltzer-Asscher (2018a), these seemingly inconsistent results observed in different studies looking at the generation of predictions highlight the fact that when forming hypotheses or interpreting results we must take into account factors that affect the nature of the predictions that are being generated (e.g. the immediacy of the predicted content, the predictive validity within the experimental context, etc.). The generation of predictions is not a single uniform process. Prediction can encompass several processes, some of which only take place under certain circumstances. It is also interesting to compare our results to those of Ness & Meltzer-Asscher (2018a). A visual inspection of the P600 pre-updating effect on the verb in both studies suggests that the magnitude of this effect was greater in the current study. We can think of two likely causes for this increased effect: i. The higher predictive validity in the current study. As explained in the Introduction, Ness & Meltzer-Asscher (2018a) proposed that the retrieval threshold is variable. If indeed pre-updating has both benefits (when

therefore test if and how the prediction, generated before the article, is altered, rather than investigating the process of generating the prediction.

successful) and costs (when unsuccessful), then different situations may vary in how beneficial it is for a comprehender to engage in pre-updating. For example, in a situation where unexpected input is often encountered even in high constraint, the costs of prediction failure may outweigh the benefit derived from facilitation of processing successfully predicted words. On the other hand, in a situation in which the input is noisy and bottom-up evidence is less reliable, the benefits of forming strong predictions may be greater. In the current experiment, the predictive validity can be considered to be 100%, since the participant provided the completions themselves, and their prediction was therefore never disconfirmed. Therefore, strong predictions are encouraged. ii. the speeded cloze task in the current study presumably provided more motivation for prediction relative to when participants only had to read for comprehension (i.e. answer a comprehension questions following the sentence). Future studies may test whether factors such as predictive validity, noisy vs. salient input, task demands, etc. in fact cause adaptation of the tendency to pre-update.

4.6) Replicability and reliability of results

In light of recent studies indicating a "replication crisis" in many research domains (e.g. Errington et al., 2014; Open Science Collaboration, 2015), the invaluable role of replication within the scientific process has become widely acknowledged (e.g. Benson, & Borrego, 2015; Bruna et al., 2017; Burman, Reed, & Alm, 2010; Fahs, Morgan, & Kalman, 2003; Sukhtankar, 2017), prompting replication attempts within the sentence processing literature (e.g. Nieuwland et. al., 2018; Nicenboim, Vasishth, & Rösler, 2019). The design of the current study enabled us to replicate the findings of Staub and colleagues (2015) regarding the effects of cloze probability and constraint on production latencies, as well as Ness & Meltzer-Asscher's (2018a) P600 pre-updating effect on the verb.

First, the behavioral results replicate the findings of Staub et al. (2015), showing the independent effects of cloze probability and constraint on production onsets. The fact that these findings were replicated in the current study using different materials and in a different language provides a very strong support not only for the replicability of the results, but also for the generalizability of the effects. Additionally, the EEG results replicate the P600 pre-updating effect found in high constraint sentences, prior to the (potentially) high cloze word, as well as the correlation between this P600 effect and WM performance (Ness & Meltzer-Asscher, 2018a). The current experiment differed from Ness & Meltzer-Asscher (2018a) in that it included a cloze (production) task, while in the previous study the sentences were only read for comprehension. Nonetheless, we view the results at the verb as a replication of the results of Ness & Meltzer-Asscher (2018a) since this point in the sentence is prior to when participants were prompted to produce a completion (in the current experiment) and the processes taking place at this point are likely similar despite the different task.

Finally, and not directly related to the main aims of the current study, having two datasets of cloze responses to the same items allowed us to assess the reliability of cloze data and the influence of the task procedure (i.e. speeded vs. offline). The results indicated a high reliability: when two different participant groups, sampled from the same general population (i.e. native Hebrew speaking, 18-40 years old, Tel-Aviv University students), provided cloze responses to the same items, highly similar values ware obtained. This held true even though the compared datasets were obtained in speeded and non-speeded tasks. This conclusion provides methodological validation for the highly prevalent use of cloze data in creating experimental materials. Additionally, as pointed out by Staub et al. (2015), it also provides justification to infer general conclusions from reaction time results in the speeded task, that apply to the common (non-speeded) task as well, since there is no ground to assume distinct generation processes in the two tasks.

5) Conclusion

The results of the current study indicate that an improbable word can, in some cases, be strongly predicted. This was shown to occur prior to when a speaker provides an improbable completion for a high constraint sentence in the cloze task. In these cases, the produced word (reflecting the participant's strongest prediction at that moment) is not only pre-activated strongly, but it can also be pre-updated (i.e. integrated into the sentence's representation in WM), which involves commitment to the prediction.

These findings support an activation race mechanism for the initiation of pre-updating, whereby multiple parallel predictions compete for activation. The activation of a given word is influenced by its probability, but also by noise or idiosyncratic influences. The most probable word, which receives the most activation from the sentence, would most often be the first to pass retrieval threshold. Nonetheless, a low cloze word would in some cases be strongly pre-activated, passing retrieval threshold prior to the high cloze competitor. Such a mechanism accounts for both the occurrence of pre-updating (demonstrated in the current study) and the generation of cloze responses (demonstrated by Staub et. al., 2015, and replicated in the current study), namely the first word to pass retrieval threshold would be integrated into the sentence's representation in WM, and/or produced as a cloze response.

6) References

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. https://doi.org/10.1016/j.jml.2012.11.001

Benson, L., & Borrego, M. (2015). The role of replication in engineering education research. Journal of Engineering Education, 104(4), 388-392. https://doi.org/10.1002/jee.20082

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. Cognition, 136, 135-149. https://doi.org/10.1016/j.cognition.2014.10.017

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. Brain research, 1446, 127-143. https://doi.org/10.1016/j.brainres.2012.01.055

Bruna, E. M., Chazdon, R., Errington, T. M., & Nosek, B. A. (2017, May 8). Reproducibility & repeatability in tropical biology: a call to repeat foundational studies. https://doi.org/10.31219/osf.io/dbpsz

Burman, L. E., Reed, W. R., & Alm, J. (2010). A call for replication studies. Public Finance Review, 38(6), 787-793. https://doi.org/10.1177/1091142110385210

Chow, W. Y., & Phillips, C. (2013). No semantic illusions in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese. Brain research, 1506, 76-93. https://doi.org/10.1016/j.brainres.2013.02.016

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why?. Current directions in psychological science, 19(1), 51-57. https://doi.org/10.1177/0963721409359277

Davis, G., & Plaisted-Grant, K. (2015). Low endogenous neural noise in autism. Autism, 19(3), 351-362. https://doi.org/10.1177/1362361314552198

Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. Brain and cognition, 135, 103569. https://doi.org/10.1016/j.bandc.2019.05.007

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. Neuropsychologia, 61, 150-162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nature Neuroscience, 8(8), 1117-1121. https://doi.org/10.1038/nn1504

Ding, J., Wang, L., & Yang, Y. (2020). The influence of emotional words on predictive processing during sentence comprehension. Language, Cognition and Neuroscience, 35(2), 151-162. https://doi.org/10.1080/23273798.2019.1628283

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. Journal of Verbal Learning and Verbal Behavior, 20(6), 641-655. https://doi.org/10.1016/S0022-5371(81)90220-6

Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., & Nosek, B. A. (2014). Science forum: An open investigation of the reproducibility of cancer biology research. Elife, 3, e04333. https://doi.org/10.7554/eLife.04333.001

Fahs, P. S., Morgan, L. L., & Kalman, M. (2003). A call for replication. Journal of Nursing Scholarship, 35(1), 67-72. https://doi.org/10.1111/j.1547-5069.2003.00067.x

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. Psychophysiology, 44(4), 491-505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. Brain research, 1146, 75-84. https://doi.org/10.1016/j.brainres.2006.06.101

Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. Brain and Language, 87(3), 345-354. https://doi.org/10.1016/S0093-934X(03)00135-4

Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: A synthesis of old and new. Current directions in psychological science, 27(6), 443-448. https://doi.org/10.1177%2F0963721418794491

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. Journal of Memory and Language, 47(2), 250-272. https://doi.org/10.1016/S0749-596X(02)00004-9

Fruchter, J., Linzen, T., Westerlund, M., & Marantz, A. (2015). Lexical preactivation in basic linguistic phrases. Journal of cognitive neuroscience, 27(10), 1912-1935. https://doi.org/10.1162/jocn_a_00822

Goldrick, M., McClain, R., Cibelli, E., Adi, Y., Gustafson, E., Moers, C., & Keshet, J. (2018). The influence of lexical selection disruptions on articulation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 45(6), 1107-1141. https://doi.org/10.1037/xlm0000633

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. Language and cognitive processes, 25(2), 149-188. https://doi.org/10.1080/01690960902965951

Green, C. (2017). Usage-based linguistics and the magic number four. Cognitive Linguistics, 28(2), 209-237. https://doi.org/10.1515/cog-2015-0112

Guo, D., Perc, M., Liu, T., & Yao, D. (2018). Functional importance of noise in neuronal information processing. Europhysics Letters, 124(5), 50001. https://doi.org/10.1209/0295-5075/124/50001

Gureckis, T. M., & Love, B. C. (2009). Learning in noise: Dynamic decision-making in a variable environment. Journal of Mathematical Psychology, 53(3), 180-193. https://doi.org/10.1016/j.jmp.2009.02.004

Hagoort, P., Brown, C. M, & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. Language and cognitive processes, 8(4), 439-483. https://doi.org/10.1080/01690969308407585

Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In The neurocognition of language (pp. 273-317). Oxford University Press.

Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. Cognitive brain research, 19(1), 59-73. https://doi.org/10.1016/j.cogbrainres.2003.10.022

Hubbard, R. J., Rommers, J., Jacobs, C. L., & Federmeier, K. D. (2019). Downstream behavioral and electrophysiological consequences of word prediction on recognition memory. Frontiers in human neuroscience, 13, 291. https://doi.org/10.3389/fnhum.2019.00291

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. Journal of Memory and Language, 86, 157-171. https://doi.org/10.1016/j.jml.2015.10.007

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. Language and cognitive processes, 15(2), 159-201. https://doi.org/10.1080/016909600386084

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. Journal of memory and language, 52(2), 205-225. https://doi.org/10.1016/j.jml.2004.10.002

Kitajo, K., Nozaki, D., Ward, L. M., & Yamamoto, Y. (2003). Behavioral stochastic resonance within the human brain. Physical Review Letters, 90(21), 218103. https://doi.org/10.1103/PhysRevLett.90.218103

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. Language, cognition and neuroscience, 31(1), 32-59. https://doi.org/10.1080/23273798.2015.1102299

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. Brain and language, 100(3), 223-237. https://doi.org/10.1016/j.bandl.2005.12.006

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A Tale of Two Positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. Journal of cognitive neuroscience, 32(1), 12-35. https://doi.org/10.1162/jocn_a_01465

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. Nature, 307(5947), 161-163. https://doi.org/10.1038/307161a0

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models. R Package Version 2.0–11. Available at: https://cran.r-project.org/package=lmerTest

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. Journal of Cognitive Neuroscience, 25(3), 484-502. https://doi.org/10.1162/jocn_a_00328

Li, X., Zhang, Y., Xia, J., & Swaab, T. Y. (2017). Internal Mechanisms underlying Anticipatory Language Processing: Evidence from Event-Related-Potentials and Neural Oscillations. Neuropsychologia, 102, 70-81. https://doi.org/10.1016/j.neuropsychologia.2017.05.017

Lowder, M. W., Choi, W., Ferreira, F., & Henderson, J. M. (2018). Lexical predictability during natural reading: Effects of surprisal and entropy reduction. Cognitive science, 42, 1166-1183. https://doi.org/10.1111/cogs.12597

Luck, S. J. (2014). An introduction to the event-related potential technique. Cambridge, MA: MIT press.

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. Journal of Memory and Language, 69(4), 574-588. https://doi.org/10.1016/j.jml.2013.08.001

McDonnell, M. D., & Ward, L. M. (2011). The benefits of noise in neural systems: bridging theory and experiment. Nature Reviews Neuroscience, 12(7), 415-425. https://doi.org/10.1038/nrn3061

McElree, B. (2001). Working memory and focal attention. Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(3), 817. https://doi.org/10.1037/0278-7393.27.3.817

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological review, 63(2), 81. https://doi.org/10.1037/h0043158

Moss, F., Ward, L. M., & Sannita, W. G. (2004). Stochastic resonance and sensory information processing: a tutorial and review of application. Clinical neurophysiology, 115(2), 267-281. https://doi.org/10.1016/j.clinph.2003.09.014

Nakamura, M. & Phillips, C. (2020, March). Why slow words sometimes finish first: Mapping cloze to activation dynamics. Poster presented at the 33rd Conference on Human Sentence Processing (CUNY), Amherst, Massachusetts. https://osf.io/2g7kr/

Ness, T. & Meltzer-Asscher, A. (2018a). Predictive pre-updating and working memory capacity: Evidence from event-related potentials. Journal of Cognitive Neuroscience, 30(12), 1916-1938. https://doi.org/10.1162/jocn_a_01322

Ness, T. & Meltzer-Asscher, A. (2018b). Lexical inhibition due to failed prediction: Behavioral evidence and ERP correlates. Journal of experimental psychology: learning, memory and cognition, 44(8), 1269-1285. https://doi.org/10.1037/xlm0000525

Nicenboim, B., Vasishth, S., & Rösler, F. (2019). Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data. https://doi.org/10.31234/osf.io/2atrh

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., ... & Mézière, D. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. ELife, 7, e33468. https://doi.org/10.7554/eLife.33468

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. https://doi.org/10.1126/science.aac4716

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. Journal of memory and language, 31(6), 785-806. https://doi.org/10.1016/0749-596X(92)90039-Z

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(4), 786. https://doi.org/10.1037/0278-7393.20.4.786

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. Cognitive Brain Research, 22(3), 407-428. https://doi.org/10.1016/j.cogbrainres.2004.09.012

Rommers, J., Dickson, D. S., Norton, J. J., Wlotko, E. W., & Federmeier, K. D. (2017). Alpha and theta band dynamics related to sentential constraint and word expectancy. Language, cognition and neuroscience, 32(5), 576-589. https://doi.org/10.1080/23273798.2016.1183799

Rommers, J., & Federmeier, K. D. (2018). Lingering expectations: A pseudo-repetition effect for words previously expected but not presented. NeuroImage, 183, 263-272. https://doi.org/10.1016/j.neuroimage.2018.08.023

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. Journal of Memory and Language, 24(2), 232-252. https://doi.org/10.1016/0749-596X(85)90026-9

Smith, N. J., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In Proceedings of the 33rd annual conference of the cognitive science society (pp. 1637–1642). Austin, TX: Cognitive Science Society. Retrieved from https://escholarship.org/uc/item/69s3541f

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. Cognition, 128(3), 302-319. https://doi.org/10.1016/j.cognition.2013.02.013

Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation durations. Psychonomic Bulletin & Review, 18(2), 371-376. https://doi.org/10.3758/s13423-010-0046-9

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. Journal of Memory and Language, 82, 1-17. https://doi.org/10.1016/j.jml.2015.02.004

Sukhtankar, S. (2017). Replications in development economics. American Economic Review, 107(5), 32-36. https://doi.org/10.1257/aer.p20171120

Szewczyk, J. M., & Wodniecka, Z. (2020). The mechanisms of prediction updating that impact the processing of upcoming word: An event-related potential study on sentence comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition. http://doi.org/10.1037/xlm0000835

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(5), 1266. https://doi.org/10.1037/0278-7393.26.5.1266 van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(3), 443. https://doi.org/10.1037/0278-7393.31.3.443

van der Groen, O., & Wenderoth, N. (2016). Transcranial random noise stimulation of visual cortex: stochastic resonance enhances central mechanisms of perception. Journal of Neuroscience, 36(19), 5289-5298. https://doi.org/10.1523/JNEUROSCI.4519-15.2016

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. International Journal of Psychophysiology, 83(2), 176-190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. Journal of cognitive neuroscience, 16(7), 1272-1288. https://doi.org/10.1162/0898929041920487

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. NeuroImage, 62(1), 356-366. https://doi.org/10.1016/j.neuroimage.2012.04.054

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. Cortex, 68, 20-32. https://doi.org/10.1016/j.cortex.2015.03.014

3 Discussion

3.1 A model of lexical prediction

Prediction during language processing has been extensively studied over the past decades. The current work focused on the specific mechanisms that underlie lexical prediction in language processing, establishing a distinction between two prediction processes: pre-activation and pre-updating. In a series of ERP and behavioral experiments, we provided evidence indicating that multiple predictions are simultaneously activated (interacting with each other), and once a highly activated prediction passes retrieval threshold it is pre-updated, i.e. integrated into the sentence representation. We showed that pre-updating is reflected in the P600 component on a word prior to the predicted word (where the prediction can be generated), and that the tendency to pre-update depends on the individual's WM abilities. We additionally provided a computational account for the pattern of influence of simultaneously pre-activated words on each other.

Taken together, these results provide support for several main aspects of my view, detailed in the Introduction, about prediction processes within the processing stages of a word (for more details also see Figure 1 and Discussion in chapter 2.1 above, Ness & Meltzer-Asscher, 2018a). Based on this view, pre-activation is 'unavoidable', it always occurs; multiple possible predictions are pre-activated to different extents, with the activation level of each word influenced by several factors such as spreading activation from prior words, the probability of the word given the input, properties of the word (e.g. frequency), etc. The current results (Ness & Meltzer-Asscher, 2021a) support the idea that multiple predictions are simultaneously pre-activated, as well as show that the activation level of a pre-activated word is influenced, among other factors, by the alternative predictions that are simultaneously activated.

Pre-updating, on the other hand, does not always occur. This process is only initiated when the activation level of a certain prediction is strong enough to pass a retrieval threshold, prior to the realization of the word in the input. Thus, pre-updating would usually be performed for highly probable predictions. However, due to the noisy nature of activations, pre-updating would occasionally be performed for improbable predictions as well. The current results provide evidence for the occurrence of pre-updating in high constraint sentences (Ness & Meltzer-Asscher, 2018b, 2021b), as well as for an activation threshold as the trigger for pre-updating (Ness & Meltzer-Asscher, 2021b).

Another important aspect of this view is that the threshold for pre-updating is variable. It can differ between individuals (due to individual differences in cognitive abilities), as well as be adapted to different situations (due to factors such as task demands, predictive validity, etc.), thus providing a mechanism to balance the benefits of pre-updating when it is successful and the costs incurred when a pre-updated prediction is disconfirmed. The need for such a mechanism stems from the assumption that disconfirmation of a pre-updated prediction incurs prediction failure costs, such as the need to inhibit the falsely predicted word in order to allow integration of another word, as shown in Ness & Meltzer-Asscher, 2018a (not included in this dissertation). The current results provide evidence that the tendency to pre-update is indeed variable between individuals. Specifically, we showed that individuals with better WM abilities (reflected in higher reading span scores) display a greater tendency to pre-update (Ness & Meltzer-Asscher, 2018b, 2021b). In another study, not included in this dissertation, we also provide evidence of adaptation of prediction strength, based on predictive validity (Ness & Meltzer-

Asscher, 2021c). In that study we manipulated the predictive validity in the experimental context (between participants), and observed that prediction failure costs decreased as the participant learned that the predictive validity in the experiment is low. This suggests that the threshold for pre-updating may be adjusted to adapt to different situations (see farther discussion below).

The view presented in this work highlights the notion that prediction is not one uniform process; rather, prediction can encompass different processes, some of which are initiated only under certain circumstances (i.e. pre-updating and inhibition). Surely, some aspects of this view remain to be tested (see below for some of the open questions), and may therefore need to be further developed or changed based on future evidence; however, a main contribution of the current view is that it explicitly discusses not only the processes involved in prediction, but also how and when each process is initiated. In proposing and detailing the different mechanisms, it provides an explanation for individual differences as well as for adaptation of prediction (via the variable threshold for pre-updating). Additionally, it helps reconcile the evidence of prediction failure costs with the evidence for prediction of specific words. As discussed in the Introduction, predicting specific words would be a "low pay-off" processing strategy, if prediction failure costs exist, since prediction of specific words are highly likely to be contradicted by the input (Forster, 1981; Jackendoff, 2002). Nonetheless, there exist both evidence for prediction of specific words (Delong, Urbach, & Kutas, 2005; Martin et al., 2013; Nieuwland et al., 2018; Nicenboim, Vasishth, & Rösler, 2020; Szewczyk & Wodniecka, 2020; van Berkum et al., 2005; Wicha, Moreno, & Kutas, 2004), and for prediction failure costs (Federmeier et al., 2007; Kuperberg, Brothers, & Wlotko, 2020; Ness & Meltzer-Asscher, 2018a). The current view helps reconcile these findings by explaining how prediction can remain a beneficial processing strategy. Namely, we propose that prediction failure costs are contingent on the occurrence of pre-updating, and pre-updating is controlled via a (variable) threshold mechanism, preventing prediction failure costs when prediction is uncertain. Hence, the distinction between pre-activation and pre-updating, and the threshold between them, allow to maintain a balance such that when it is not likely to be beneficial to commit to a specific prediction, due to a high probability of failure (e.g. in a low constraint context, and/or when predictive validity is low), prediction will only manifest in graded pre-activation, and will not incur failure costs; and only when committing to a specific prediction has a high probability of success (e.g. in a high constraint context, when predictive validity is not low), such commitment is engaged in, i.e. pre-updating occurs. This way, the generation of specific word predictions does not entail a frequent occurrence of prediction failure costs.

3.2 Open questions and future directions

3.2.1 Pre-updating and inhibition

As mentioned above, a main assumption of my view is that disconfirmation of a pre-updated prediction incurs prediction failure costs, that are not incurred for the processing of an unexpected word if no prediction was pre-updated in the first place. Specifically, in Ness & Meltzer-Asscher (2018a, not included in this dissertation) we suggested that these costs stem from a need to inhibit the falsely predicted word (but see also Kuperberg, Brothers, and Wlotko, 2020, who argue that these costs reflect suppression of an event representation rather than inhibition at the word level). In that study, we looked at high constraint sentences, ending with either the predictable word, a congruent unexpected word, or an anomalous word. Using CMLP, we showed that the highly predicted word is strongly activated prior to its anticipated appearance, but is then inhibited if a congruent unexpected word appears instead (but see Federmeier & Rommers, 2018, for indication that disconfirmed predictions may not be fully inhibited). Interestingly, this inhibition is not observed if an anomalous word appears. Since inhibition of the falsely predicted word does not take place when encountering an anomaly, we have argued that this inhibition is required particularly to enable integration of the word that actually appeared in the input (see Kutas 1993, and Ness & Meltzer-Asscher, 2018a, for further discussion of this claim).

The assumption that pre-updating involves integration of the prediction into the sentence representation makes the association between pre-updating and inhibition very intuitive, i.e. the integration of a prediction into the sentence representation is what causes the need to inhibit it if another word needs to be integrated instead. However, thus far we have not provided direct evidence for this relation. Namely, we demonstrated inhibition under conditions in which pre-updating is likely to occur, i.e. the behavioral inhibition and the f-PNP ERP component were observed in high constraint sentences (Ness & Meltzer-Asscher, 2018a), and prediction failure costs were greater when predictive validity in the experimental context was high (Ness & Meltzer-Asscher, 2021c), but we have not directly shown that inhibition occurs only when pre-updating has previously occurred. Obtaining direct evidence for the relation between pre-updating and inhibition poses some methodological challenges, since it is not sufficient to demonstrate that, across participants and/or trials, both pre-updating and inhibition occur under the same circumstances (since these circumstances may independently cause both pre-updating and inhibition, without a direct link between the two processes). Rather, a more direct evidence would require a demonstration that when comparing two similar trials, which differ only in whether the specific participant in that moment in time preformed pre-updating or not, inhibition only occurs if pre-updating has occurred (in that specific trial), but not if pre-updating has not occurred.

3.2.2 Individual differences in prediction processes

The current work provided indication that individual differences in WM abilities affect the tendency to pre-update, by demonstrating correlation between the pre-updating P600 effect and reading span scores (Ness & Meltzer-Asscher, 2018b, 2021b). Notably, reading span scores are a relatively non-specific measure, which may be influenced by multiple cognitive constructs (see e.g. Daneman & Hannon, 2007; Miyake, 2001). Moreover, additional cognitive abilities that are not necessarily reflected in reading span scores may affect the tendency to pre-update. For example, it is likely that the tendency to pre-update is influenced not only by resources available to allow pre-updating to occur, but also from the ability to successfully recover when a pre-updated prediction is disconfirmed. Hence, it would be expected that individuals with better ability to perform inhibition would have a greater tendency to pre-update (as the 'risk' is smaller for them). Future work can focus on identifying additional factors that modulate individual differences in the tendency to pre-update, by employing a wider variety of tasks. Potentially, the tendency to pre-update may be correlated with individual differences in the ability to generate accurate predictions (e.g. language proficiency, statistical learning abilities), the ability to perform pre-updating (e.g. WM-related abilities, general processing speed), and the ability to recover from a disconfirmed pre-updated prediction (e.g. the ability to inhibit irrelevant distractors).

3.2.3 Adaptation of the retrieval threshold

In a recent study we have shown that comprehenders can adapt their prediction processes to different situations, such that prediction failure costs are reduced if the participant estimates that the predictive validity in the experimental context is low (Ness & Meltzer-Asscher, 2021c). We argue that a plausible mechanism for such adaptation is the adjustment of the threshold for pre-updating. Namely, comprehenders alleviate prediction failure costs when prediction validity is low by raising the threshold, thus avoiding pre-updating, which prevents the prediction failure costs associated with disconfirmation of pre-updated predictions. Importantly, since the experiments reported in Ness & Meltzer-Asscher (2021c) are behavioral, only the effect of prediction failure is measured (prolonged reading times for disconfirmed predictions), with no direct indication of pre-updating (as would be reflected by the P600 pre-updating effect). In order to support the aforementioned adaptation mechanism, there is a need to conduct a similar study but with the addition of ERP recording, which would allow to examine whether the occurrence of pre-updating (reflected in the P600 effect) is indeed reduced when the participant's estimate of predictive validity is lowered.

More broadly, if the threshold for pre-updating is indeed adjustable, there may be additional factors (other than predictive validity) which can trigger its adjustment. Since the threshold is hypothesized to keep the balance between the benefits of successful prediction and the costs of prediction failure, identifying the specific factors that trigger adaptation may inform us about what makes prediction beneficial. Several potential benefits have been suggested for the use of prediction as a language processing strategy (see Huettig, 2015, for discussion of motivations for prediction). For example, prediction may be helpful in reducing the ambiguity that exists in most linguistic input, either due to semantically/grammatically ambiguous utterances or due to perceptual ambiguity (e.g. arising from noisy input and production variation), by constraining the interpretation of the input to more probable meanings/representations. Additionally, prediction has also been argued to enable coordinated 'turn taking' during dialogue. The specific benefits comprehenders derive from prediction determine what factors should influence how beneficial prediction is in different situations. For example, if prediction is indeed helpful in disambiguating perceptually ambiguous input, then it may be more beneficial to generate strong predictions in a noisy environment than in a quiet one; if prediction is needed to coordinate 'turn taking', it may be more beneficial to generate strong predictions during a conversation than during passive listening (e.g. listening to a lecture or watching a movie). Thus, further work can examine additional factors, such as noise and task demands, which may trigger adaptation of the threshold for pre-updating.

3.2.4 Prediction processes in neurodiverse populations

Several studies have shown decreased predictive abilities in language processing in older adults (e.g. Dave et al., 2018; DeLong et al., 2012; Wlotko, Federmeier, & Kutas, 2012), and in different neurocognitive disorders (e.g. in aphasia: Hanne et al., 2015; Warren, Dickey, & Lei, 2016). A better understanding of the specific mechanisms that underlie prediction can help provide a more accurate account of the impaired processes in each population. For example, recent studies on prediction in older adults indicate differential influence of age on N400 effects and on PNP effects (Dave et al., 2018; Wlotko, Federmeier, & Kutas, 2012). This may suggest that pre-activation and pre-updating are differentially affected by aging, but further research is needed in order to support this claim. Similarly, further research is needed in order to identify how pre-activation and pre-updating are affected in different neurocognitive disorders.

4 References for Introduction and Discussion

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419–439.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264.

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57(4), 502–518.

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, 28(4), 525–560.

Boland, J. E. (2005). Visual arguments. *Cognition*, 95(3), 237-274.

Brothers, T., Dave, S., Hoversten, L. J., Traxler, M. J., and Swaab, T. Y. (2019). Flexible predictions during listening comprehension: speaker reliability affects anticipatory processes. *Neuropsychologia* 135: 107225.

Brothers, T., Swaab, T. Y., and Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, 93, 203–216.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, 136, 135-149.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive science*, 41, 1318-1352.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain research*, 1446, 127-143.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of cognitive neuroscience*, 5(1), 34-44.

Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in human neuroscience*, 4, 25.

Clark, A. (2015). Radical predictive processing. *The Southern Journal of Philosophy*, 53, 3-27.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3), 687–696.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological review*, 119(2), 417.

Chow, W. Y., & Phillips, C. (2013). No semantic illusions in the “Semantic P600” phenomenon: ERP evidence from Mandarin Chinese. *Brain research*, 1506, 76-93.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, 1084(1), 89-103.

Daneman, M., & Hannon, B. (2007). What do working memory span tasks like reading span really measure. In N. Osaka, R. H. Logie, & M. D'Esposito (Eds) *The cognitive neuroscience of working memory* (pp. 21-42). Oxford University Press.

Dave, S., Brothers, T. A., Traxler, M. J., Ferreira, F., Henderson, J. M., & Swaab, T. Y. (2018). Electrophysiological evidence for preserved primacy of lexical prediction in aging. *Neuropsychologia*, 117, 135-147.

DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and language*, 121(3), 226-239.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8 (8), 1117–1121.

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science*, 21(5), 629–634.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20(6), 641-655.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, USA, 103(32), 12203–12208.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491-505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4), 469–495.

Federmeier, K.D., Laszlo, S., 2009. Time for meaning: electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychology of learning and motivation*. 51, 1–44

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain research*, 1146, 75-84.

Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. *Brain and Language*, 87(3), 345-354.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, 47(2), 250-272.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology*, 33(4), 465-495.

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language*, 95, 200-214.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37(1), 58–93.

Gibson, E., & Wu, H. H. I. (2013). Processing Chinese relative clauses in context. *Language and Cognitive Processes*, 28 (1–2), 125–155.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and cognitive processes*, 25(2), 149-188.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.) *The cognitive neurosciences*, 4th ed. (pp. 819-836). MIT press.

Hagoort, P., Brown, C. M, & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and cognitive processes*, 8(4), 439-483.

Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In *The neurocognition of language* (pp. 273-317). Oxford University Press.

Hanne, S., Burchert, F., De Bleser, R., & Vasishth, S. (2015). Sentence comprehension and morphological cues in aphasia: What eye-tracking reveals about integration and prediction. *Journal of Neurolinguistics*, 34, 83-111.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, 48(2), 281–303.

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in neuroscience*, 8, 150.

Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive brain research*, 19(1), 59-73.

Hohwy, J. (2018). The predictive processing hypothesis. In A., Newen, L., De Bruin, & S., Gallagher (Eds.) *The Oxford handbook of 4E cognition* (pp. 129-145), Oxford University Press.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain research*, 1626, 118-135.

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 645.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, 86, 157-171.

Jackendoff, R. 2002. Foundations of language: brain, meaning, grammar, evolution. New York: Oxford University Press.

Jackson, S. A., Kleitman, S., Howie, P., & Stankov, L. (2016). Cognitive abilities, monitoring confidence, and control thresholds explain individual differences in heuristics and biases. *Frontiers in Psychology*, 7, 1559.

James, W. (1890). The Principles of Psychology, Vol. 1. New York: Dover Publications.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and cognitive processes*, 15(2), 159-201.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The timecourse of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of memory and language*, 52(2), 205-225.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and cognitive processes*, 8(4), 533-572.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in cognitive sciences*, 4(12), 463-470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, 62, 621-647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161–163.

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In: S. Kornblum, & J. Requin (Eds.), *Preparatory States and Processes* (pp. 217-237). Hillsdale, NJ: Lawrence Erlbaum.

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A Tale of Two Positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of cognitive neuroscience*, 32(1), 12-35.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, cognition and neuroscience*, 31(1), 32-59.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and language*, 100(3), 223-237.

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: an ERP study. *Journal of Cognitive Neuroscience*, 23(5), 1230–1246.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, 25(3), 484-502.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, 88, 22-60.

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, 69(4), 574-588. https://doi.org/10.1016/j.jml.2013.08.001

Miyake, A. (2001). Individual differences in working memory: Introduction to the special section. *Journal of Experimental Psychology: General*, 130(2), 163.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 106(3), 226.

Ness, T., & Meltzer-Asscher, A. (2018a). Lexical inhibition due to failed prediction: Behavioral evidence and ERP correlates. *Journal of experimental psychology: Learning, memory, and cognition*, 44(8), 1269.

Ness, T. & Meltzer-Asscher, A. (2018b). Predictive pre-updating and working memory capacity: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, 30(12), 1916-1938.

Ness, T. & Meltzer-Asscher, A. (2021a). Love thy neighbor: facilitation and competition between parallel predictions. *Cognition*, 207, 104509.

Ness, T. & Meltzer-Asscher, A. (2021b). From pre-activation to pre-updating: A threshold mechanism for commitment to strong predictions. *Psychophysiology*, 58(5), e13797.

Ness, T. & Meltzer-Asscher, A. (2021c). Rational adaptation in lexical prediction: The influence of prediction strength. *Frontiers in Psychology*, 12, 1166.

Nicenboim, B., Vasishth, S., & Rösler, F. (2020). Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia*, 142, 107427.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., ... & Mézière, D. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, 7, e33468.

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., ... & Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society* B, 375(1791), 20180522.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, 31(6), 785-806.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 786.

Paczynski, M., & Kuperberg, G. R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb argument processing. *Language and Cognitive Processes*, 26(9), 1402–1456.

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/ state knowledge and animacy selection restrictions. *Journal of Memory and Language*, 67(4), 426–448.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22(3), 407-428.

Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. *Cognition*, 118(3), 339–358.

Rommers, J., & Federmeier, K. D. (2018). Lingering expectations: A pseudo-repetition effect for words previously expected but not presented. *NeuroImage*, 183, 263-272.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(2), 344.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, 24(2), 232-252.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, 112(1), 1.

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, 82, 1-17.

Szewczyk, J. M., & Wodniecka, Z. (2020). The mechanisms of prediction updating that impact the processing of upcoming word: An event-related potential study on sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

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.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, 83(3), 382-392.

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1266.

Van Berkum, J. J. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In *Semantics and pragmatics: From experiment to theory* (pp. 276-316). Palgrave Macmillan.

van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443.

Van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive brain research*, 22(2), 241-255.

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes*, 8(4), 485-531.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 394.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brainpotentials. *Memory & Cognition*, 18(4), 380-393.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open-and closed-class words. *Memory & Cognition*, 19(1), 95-112.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176-190.

Warren, T., Dickey, M. W., & Lei, C. M. (2016). Structural prediction in aphasia: Evidence from either. *Journal of neurolinguistics*, 39, 38-48.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of cognitive neuroscience*, 16(7), 1272-1288.

Wilson, M. P., & Garnsey, S. M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language*, 60(3), 368–392.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, 62(1), 356-366.

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: age-related differences in the use of sentential context. *Psychology and aging*, 27(4), 975.

Xiang, M., & Kuperberg, G. R. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 30(6), 648–672.

תקציר

חיזוי בעת עיבוד שפה היווה מושא למחקר רב במהלך העשורים האחרונים, עם מחקרים רבים המתמקדים במנגנונים המעורבים בחיזוי, המאפיינים שלהם, והאופן בו הם באים לידי ביטוי במדדים התנהגותיים ומוחיים. העבודה הנוכחית מתמקדת בהבחנה בין שני תהליכי חיזוי: "אקטיבציה מוקדמת", איקטוב של מידע לקסיקאלי/סמנטי המאוחסן בזיכרון ארוך-טווח, ו-"עדכון מוקדם", עדכון ייצוג המשפט הנבנה בזיכרון העבודה, לכלול את תוכן החיזוי.

בעבודה זו אני מציעה ובוחנת מודל עבור פעולת תהליכים אלו והשתלבותם בשלבי העיבוד של מילה במשפט. על פי מודל זה, חיזויים שונים מאוקטבים במקביל ובאופן מדורג (בהתאם לעוצמת החיזוי של כל מילה). עדכון מוקדם מתרחש רק אם חיזוי כלשהו מאוקטב באופן חזק מאוד. כלומר, רק אם מידת האקטיבציה של חיזוי מסויים עוברת סף שליפה, מתבצע עדכון מוקדם, והמילה שהייתה מאוקטבת מאוד עוברת אינטגרציה אל תוך ייצוג המשפט כאשר החיזוי מתרחש, במקום כאשר המילה מופיעה בקלט. אם חיזוי שעבר עדכון מוקדם מתברר כשגוי, שגיאת החיזוי תגרור עלות עיבודית (לצורך ביצוע אינהיביציה של המילה השגויה שעודכנה). היבט חשוב של מודל זה הוא שהסף לעדכון מוקדם אינו קבוע. סף זה יכול להיות שונה בין אנשים שונים (בשל הבדלים בינאישיים במאפיינים כגון יכולות זיכרון עבודה). בנוסף, הסף יכול לעבור התאמה לסיטואציות שונות (בשל מאפיינים כמו מהימנות הניבוי בסיטואציה, רמות הרעש בקלט, וצרכי המטלה). בכך סף זה שולט בנטייה לבצע עדכון מוקדם, על מנת לאזן בין הפסיליטציה המתקבלת בעת חיזוי נכון לבין העלות העיבודית של חיזוי שגוי.

עבודה זו מורכבת משלושה מאמרים, המדווחים על סדרת ניסויים התנהגותיים ואלקטרופיזיולוגיים (event-related potentials, ERP) שמטרתם לבחון את האספקטים העיקריים של המודל שהוצע לעיל. במאמר הראשון מוצגת עדות אלקטרופיזיולוגית לקיומו של תהליך העדכון המוקדם, אשר מתבטאת כאמפליטודה גבוהה של קומפוננטת ה P600 -במשפטים המייצרים חיזוי חזק (לעומת משפטים שאינם מייצרים חיזוי חזק), על הפועל המוביל למילה מאוד צפויה. בנוסף אנו מראות שאפקט זה הינו בקורלציה חיובית עם ציוני הנבדקים במבדק זיכרון עבודה (reading span), תוצאה המעידה כי הנטייה לבצע עדכון מוקדם משתנה בין נבדקים בהתאם ליכולות זיכרון העבודה.

במאמר השני אנו בוחנות את מהירות תחילת ההפקה במטלת השלמת משפטים מזורזת (speeded cloze task), כלומר, מטלת השלמת משפטים שבה הנבדק נדרש להפיק בקול מילה המשלימה את המשפט במהירות האפשרית. אנו מראות שמהירות תחילת ההפקה של ההשלמה הנפוצה ביותר עבור כל משפט מושפעת מעוצמת הניבוי והקשר הסמנטי של מתחרה שאינו מופק, המילה השניה הצפויה ביותר. תוצאות אלה תומכות בכך שחיזויים שונים מאוקטבים במקביל, ומראות כי רמת האקטיבציה של מילה חזויה מושפעת מהחיזויים האלטרנטיביים הקיימים. במאמר זה אנחנו גם מספקות מודל חישובי אשר ממדל את מהירות תחילת ההפקה במטלת השלמת משפטים מזורזת, ע"י שינוי והרחבה של מודל האינטראקציה אקטיבציה ותחרות (interactive activation and competition model, IAC) של Chen ו- Mirman (2012).

במאמר השלישי אנו חוקרות את הנסיבות בהן עדכון מוקדם מתבצע, באמצעות שימוש במטלת השלמת משפטים מזורזת תוך הקלטת ERP על הפועל המופיע לפני הפקת ההשלמה למשפט. שיטה זו איפשרה לנו לנתח את נתוני ה- ERP בהתבסס על המילה הספציפית שהפיק/ה כל נבדק/ת בכל צעד ניסוי, אשר משקפת את החיזוי החזק ביותר עבור הנבדק/ת הספציפי/ת ברגע המסויים. התוצאות שיחזרו את הממצא של אמפליטודת P600 גבוהה במשפטים המייצרים חיזוי חזק (לעומת משפטים שאינם מייצרים חיזוי חזק), על הפועל המוביל למילה מאוד צפויה. יתרה מכך, אפקט זה התקיים במשפטים המייצרים חיזוי חזק (לעומת משפטים שאינם מייצרים חיזוי חזק), אפילו כאשר החיזוי החזק ביותר עבור הנבדק/ת בצעד ניסוי מסויים (המשתקף במילה שהופקה) היה של מילה עם סבירות נמוכה. תוצאות אלה תומכות במרוץ אקטיבציה רועש אל עבר סף שליפה כמנגנון המפעיל את תהליך העדכון המוקדם.

בכללן, תוצאות אלה מקדמות את ההבנה של המנגנונים הספציפיים המעורבים בחיזוי בזמן עיבוד שפה, מדגישות את הרעיון שחיזוי אינו תהליך אחיד בודד, ומקדמות מודל מפורט יותר של מנגנוני החיזוי.

TEL AVIV UNIVERSITY אוניברסיטת תל אביב
בית הספר סגול למדעי המוח

מנגנוני חיזוי בעיבוד משפטים: לא רק אקטיבציה

חיבור לשם קבלת תואר
"דוקטור לפילוסופיה"

מאת
טל נס

החיבור בוצע בהנחייתה של
פרופסור איה מלצר-אשר

הוגש לסנאט אוניברסיטת תל-אביב
יולי, 2021