LANGUAGE, COGNITION AND NEUROSCIENCE
2024, VOL. 39, NO. 5, 632–656
https://doi.org/10.1080/23273798.2024.2345302

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS Check for updates

Testing two-step models of negative quantification using a novel machine learning analysis of EEG

S. Ramotowska [a], K. Archambeau [b], P. Augurzky [c], F. Schlotterbeck [d], H.S. Berberyan [e], L. Van Maanen [f] and J. Szymanik [g]

[a]Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, The Netherlands; [b]Université Libre de Bruxelles, Bruxelles, Belgium; [c]Department of Psychology, Universität Tübingen, Tübingen, Germany; [d]Institute of German Language and Literatures, Universität Tübingen, Tübingen, Germany; [e]Bernoulli Institute, University of Groningen, Groningen, The Netherlands; [f]Department of Experimental Psychology & Helmholtz Institute, Utrecht University, Utrecht, The Netherlands; [g]Center for Mind/Brain Sciences and Dept. of Information Engineering and Computer Science, University of Trento, Trento TN, Italy

ABSTRACT
The sentences "*More than half* of the students passed the exam" and "*Fewer than half* of the students failed the exam" describe the same set of situations, and yet the former results in shorter reaction times in verification tasks. The two-step model explains this result by postulating that negative quantifiers contain hidden negation, which involves an extra processing stage. To test this theory, we applied a novel EEG analysis technique focused on detecting cognitive stages (HsMM-MVPA) to data from a picture-sentence verification task. We estimated the number of processing stages during reading and verification of quantified sentences (e.g. "*Fewer than half* of the dots are blue") that followed the presentation of pictures containing coloured geometric shapes. We did not find evidence for an extra step during the verification of sentences with *fewer than half*. We provide an alternative interpretation of our results in line with an expectation-based pragmatic account.

ARTICLE HISTORY
Received 7 July 2023
Accepted 5 April 2024

KEYWORDS
Two-step model; electroencephalography; polarity effect; quantifiers; hidden semi-Markov model multivariate pattern analysis

1. Introduction

In the 1960s, studies first showed that sentences with negation (e.g. "Nine is not an even number") take longer to process than affirmatives (e.g. "Nine is an odd number") (Wason, 1961). However, this effect cannot be straightforwardly attributed to the meaning of negation itself. Amongst other reasons, this is because explicit negation lengthens the sentence: the longer the sentence, the more complex it is, and, therefore, the longer it takes to process (see Grodzinsky et al., 2020, for methodological discussion). To avoid this confound, Just and Carpenter (1971) tested three types of negation: explicit syntactic negatives (e.g. *none*), implicit syntactic negatives (e.g. *few*), and semantic negatives (e.g. *a minority*). They found that participants verified all types of negatives longer than affirmatives. Because the sentences with implicit syntactic and semantic negatives and affirmatives were the same length, this study confirmed that the processing difficulties related to negation are not just a function of the length of the sentence but are inherent to negation. This highly replicable effect is called the polarity effect, a general linguistic phenomenon of negative expressions (including sentential negation) being more difficult to process than their affirmative counterparts (Deschamps et al., 2015; Just & Carpenter, 1971, see Clark, 1976 for review).

Several theoretical proposals aimed to explain this effect (e.g. Clark & Chase, 1972; Grodzinsky et al., 2018; Kaup et al., 2006). In this paper, we discuss and test one of the general approaches, namely the two-step model (see Clark, 1976, for review). We refer to the two-step model as a class of models that share a common assumption: they postulate that negation and negative expressions involve an extra processing step. To control for the confound caused by the explicit negation (the length of the sentence), we investigated the polarity effect by testing a pair of quantifiers that were comparable in length: positive (*more than half*) and negative (*fewer than half*).

The two-step model was inspired by studies on sentential negation (Clark, 1976; Clark & Chase, 1972; Kaup

CONTACT S. Ramotowska ramotowska.or.s@gmail.com, sonia.ramotowska@uva.nl
Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/23273798.2024.2345302.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

et al., 2006). It is also well-grounded in the semantic analysis of negation and negative expressions (Grodzinsky et al., 2018). It appeals to the idea that a sentence is processed in a sequence of stages. These processing stages correspond to the mental operations of building the representation of the sentence. The more complex the sentence, the more operations it involves. The main assumption of the two-step model is that negative expressions (e.g. the quantifier *fewer than half*) involve an additional mental operation. This extra mental operation explains the longer latency of sentence processing involving negative expressions. It should therefore be reflected in reaction time differences, namely, it should take longer to process negatives than affirmatives. This prediction bore out in behavioural studies on explicit negation, expressed in English by *no*, *not*, *it is not true that* (Just & Carpenter, 1971); and implicit negation, expressed by negative quantifiers (e.g. *few, fewer than half*, Schlotterbeck et al., 2020), adjectives (e.g. *short*, Tucker et al., 2018), or location words (e.g. *below*, Clark & Chase, 1972).

Thus far, the two-step model has mostly been tested indirectly. The experimental studies measuring mean reaction times (e.g. Clark & Chase, 1972; Just & Carpenter, 1971; Kaup et al., 2006) or event-related potentials (ERPs) (e.g. Farshchi et al., 2020; Fischler et al., 1983) can only indirectly support the two-step model by postulating a linking assumption between the model's predictions and the data pattern. For example, in the reaction time experiments, it is assumed that the difference in mean reaction times between experimental conditions reflects the extra processing step. Similarly, electroencephalographic (EEG) experiments assume that the difference in the ERP components is due to the extra processing step.

The idea that the upcoming information (for example, a sentence) is processed in a series of cognitive stages has a long tradition not only in linguistics but also in experimental psychology (Donders, 1969; Sternberg, 1969). The processing stages postulated in cognitive models are also reflected in the stages of processing in the brain (Zylberberg et al., 2011). Recent advancements in computational modelling allow us to directly estimate the number of processing stages in simple cognitive tasks (Anderson et al., 2016). This approach (HsMM-MVPA) makes use of the multivariate nature of EEG to estimate the most likely onsets of cognitive processing stages. As a result, the method also identifies the most likely number of processing stages for a particular task or experimental condition. This way, we can directly test the two-step model, as it predicts that the number of processing stages is related to the polarity effect.

In the next sections, we will explain the key concepts of the two-step model and present experimental findings that indirectly support its predictions. Then, we will point out the limitations of these studies and show how we can directly test the two-step model by using the Hidden semi-Markov Model Multivariate Pattern Analysis (HsMM-MVPA).

1.1. Two-step models

Broadly speaking, the two-step models of (explicit or implicit) negation assume that an additional processing step related to the processing of negation could be mapped onto cognitive stages. This family of models roots from studies on sentential negation, but it can be extended to other types of negatives. Two-step models postulate two sources of processing difficulties. The first one is at the level of mental representation and refers to the complexity of this representation (Agmon et al., 2022; Clark & Chase, 1972; Grodzinsky et al., 2018; Kaup et al., 2006). The second source lies in the verification procedure and steps involved in the computation of a sentence truth value (Agmon et al., 2022; Clark & Chase, 1972; Grodzinsky et al., 2018).

Various semantic analyses explain the source of the representational complexity of negatives. According to the two-step simulation hypothesis, the representation of sentences negation like "A is not above B" contains the positive proposition "A is above B", called the to-be-negated state (Kaup et al., 2006). To access the representation of the actual state of affairs, firstly, participants have to represent the to-be-negated sentence and mentally tag it as false. The simulation account explicitly postulates an extra step in the processing of the negated sentence. The model found supportive evidence coming from reaction time data from picture-sentence verification experiments (Kaup et al., 2006, 2007). For example, Kaup et al. (2006) showed that a delay of 1500 ms of picture presentation is sufficient for participants to shift their attention from the representation of the to-be-negated state and focus on the actual state of affairs.

Already Clark and Chase (1972), (see also Clark, 1976) observed that implicit negations (e.g. locatives *below*) take longer to process than positives (*above*). Their model of sentence negation processing, the so-called "true" model of negation (Clark & Chase, 1972, see also Clark, 1976, also known as the schema-plus-tag model), included a parameter relating to this difference in the encoding time.

Agmon et al. (2019) proposed that the property of negative polarity adds complexity to the mental representation of negatives. According to their proposal,

positive expressions always have denotations above a certain reference point on a mental scale. For example, *more than half* means *more than the threshold of half*, *tall* means *more than a certain height*, and *above* means *more up than the reference point* see similar argument in Clark (1976). Negatives, in turn, refer to the opposite direction. They are cognitively costly because they reverse a natural order on the scale. As a result, negative expressions have a more complex representation because they contain hidden negation and involve a computation of *less than* scale reversal operation.

Furthermore, in the realm of quantifiers, Grodzinsky et al. (2018) proposed that representational complexity is related to the number of downward entitling operators. Downward and upward entailment refers to the opposite entailment patterns. For sets A and A' if $A \subseteq A'$, then a quantifier Q is upward entailing if $Q(A) \subseteq Q(A')$ and downward entailing if $Q(A') \subseteq Q(A)$. For example, the sentence "*More than half* of the men run fast" entails that "*More than half* of the men run," while the sentence "*Fewer than half* of the men run" entails that "*Fewer than half* of the men run fast ". According to Grodzinsky et al. (2018), the comparative *more* is represented as *many* + *er*, while *fewer* is represented as *little* + *many* + *er*, where *little* is an extra downward entailing operator not present in *more* (cf. Heim, 2006).[1] Recently, Agmon et al. (2022) suggested that downward monotonicity might contribute to the polarity effect by increasing working memory load. When verifying negative quantifiers, participants have to retrieve a non-default entailment pattern, which in turn results in increasing the processing cost.

Concerning the verification of negative expressions, already early models of negation (Clark & Chase, 1972; Young & Chase, 1971b) made different predictions about the interaction between the sentence polarity and the sentence truth value. The "true" model of negation predicts the interaction, namely that affirmative sentences should be verified faster when they are true than when they are false, and the opposite pattern of reaction times for negative sentences.[2] The "conversion" model, in contrast, assumes that a negative sentence can be converted into an affirmative sentence and verified after the conversion. This model postulates longer reaction times for all negative sentences (see Clark, 1976, for a detailed description of this model) and predicts only the main effects of truth value and negation, but no interaction.[3] While some studies support the "conversion" model (Young & Chase, 1971a as cited in Clark, 1976, cf. Wason, 1961), the model's application is limited to tasks with two contradictory predicates (e.g. *odd number* vs. *even number*) where one is a negation of another (e.g. *odd number* means *not even number*). Similar to sentence negation, implicit negatives involve a longer verification procedure. Grodzinsky et al. (2018) postulated that quantifiers, in addition to representational complexity, require a more complex verification procedure (verification complexity, cf. Barwise & Cooper, 1981). The predictions of this model bore out in recent reaction time experiments (Agmon et al., 2022).

To conclude, the two-step models of sentence negation and various implicit negations make two main predictions concerning the additional step of processing. Firstly, they predict that a more complex semantic representation of negatives involves an additional mental operation. Secondly, they predict that the procedure of truth value computation might involve more steps in the case of negative than positive expressions.

1.2. Electroencephalographic evidence for two-step models

Besides the evidence from reaction time experiments, the two-step model is also supported by electroencephalographic (EEG) findings. Classical EEG studies on language processing use the event-related potential (ERP) technique, which involves averaging the signal over trials and participants. Two components are particularly interesting for language processing – the N400 and P600 (Delogu et al., 2019). The N400 component is sensitive to semantic mismatch and incongruency (Kutas & Hillyard, 1980), as well as to world knowledge, discourse, cloze probability, and non-linguistic meaning processing (see Kutas & Federmeier, 2011, for review). It is a signature of the lexical retrieval processes (Delogu et al., 2019). The P600, in turn, was first linked to syntactic processing (Hagoort et al., 1993; Kaan et al., 2000; Osterhout & Holcomb, 1992), but is also related to semantic integration (Brouwer et al., 2017, 2012).

Early EEG evidence for the two-step processing of negation comes from the phenomenon called negation-blind N400 (Fischler et al., 1983). A sentence like "A dog is a fish" is false and semantically incongruent. It should, therefore, elicit an N400 on the final word of the sentence (*fish*). Fischler et al. (1983) showed that the N400 was induced not only by false sentences like "A dog is a fish," but also by true negative sentences like "A dog is *not* a fish," which is a correct and semantically congruent sentence. The amplitude difference and latency of the N400 were comparable between affirmative and negative sentences. The lack of an N400 reduction in the presence of negation was interpreted as evidence of a delay in processing. Palaz et al. (2020)

showed a similar result in a more pragmatically felicitous context.

In another study, Dudschig and Kaup (2018) used the lateralised readiness potential (LRP) and showed that the to-be-negated information is initially activated. They argued that the clash between negated information and the actual state of the world is processed similarly to a conflict in conflict-monitoring tasks (Botvinick et al., 2001; Van Maanen & Van Rijn, 2010; Van Maanen et al., 2012). The idea that negation requires switching between two mental representations was further supported by the EEG signatures of response inhibition in negation processing (Beltrán et al., 2019). These findings support the idea that the explicit negation is represented in two steps and that additional cognitive resources are needed to choose between the representations.

A few studies (Augurzky et al., 2020; Nieuwland, 2016; Urbach et al., 2015; Urbach & Kutas, 2010) tested the online processing of negative and positive quantifiers. For example, Urbach and Kutas (2010) manipulated the lexical-semantic associations between quantifiers (*most*, *few*), adverbs (*often*, *rarely*) and nouns to create typical and atypical sentences. They expected to find a cross-over interaction between quantifier/adverb and typicality, as reflected by an N400 component. What they found, however, was an asymmetry in N400 amplitude for positive vs. negative quantifiers. The N400 effect followed the predicted patterns only for positive expressions. Moreover, they found that the prefrontal positivity in atypical sentences was more pronounced for negative expressions, suggesting that negative expressions require additional processing compared to positive ones. In a follow-up experiment, Urbach et al. (2015) demonstrated that in a pragmatically appropriate discourse context, the N400 pattern follows the expected full cross-over interaction pattern. Further studies (Nieuwland & Kuperberg, 2008; Urbach et al., 2015) demonstrated that negative expressions can be processed easier in a pragmatically felicitous context, while others (Orenes et al., 2016) showed that sentences with negation are still processed slower than affirmative sentences.

Further evidence for a delay in the processing of negative quantifiers comes from a picture-sentence verification task (Augurzky et al., 2020). Because previous studies (Nieuwland & Kuperberg, 2008; Urbach & Kutas, 2010) showed that discourse information can affect the processing of negative quantifiers or negation, Augurzky et al. (2020) presented participants contexts given as a picture instead of a sentence. Pictures, in contrast to world-knowledge-based sentences, were equally informative for all quantifiers in the experiment. In addition, by presenting quantified sentences in the context of a picture, Augurzky et al. (2020) were able to control for lexical associations as a potential confounding factor.

Moreover, previous experiments (Urbach et al., 2015; Urbach & Kutas, 2010) tested the polarity effect by comparing quantifiers, for example *few* and *most*. However, in addition to polarity, this pair of quantifiers differs also in other semantic properties. For example, *most* is a superlative quantifier (Hackl, 2009), while *few* is not. For this reason, Augurzky et al. (2020) chose the quantifiers *more than half* and *fewer than half* with highly comparable semantic properties. They tested the online verification of sentences such as "*More than half*/*Fewer than half* of the dots are yellow ", and found a contrast in the N400 measured on the adjective onset for false vs. true sentences when the quantifier was *more than half* and no effect for *fewer than half*. They interpreted this finding as evidence for the delay in processing negative quantifiers.

In addition to the analysis of the N400 effect, Augurzky et al. (2020) conducted an exploratory analysis and found a greater late positivity activation for *fewer than half* than for *more than half*. They measured the ERPs from the onset of *more/fewer*, which was presented separately from *than half*. The authors discussed two possible interpretations of this finding. According to the first one, processing of *fewer than half* is more cognitively costly than *more than half*, and the late positivity reflects an increase in attentional demands. According to the second interpretation, the positive component is related to a contextual update. The participants encoded the picture in terms of the larger proportion, and as soon as they processed *fewer than half*, they had to revise their current discourse model. Both explanations suggest that the origin of the N400 delay effect could be traced back to difficulties in processing the negative quantifier.

The analysis of the above-mentioned late positivity was exploratory and did not directly show that the difficulties in processing negative quantifiers were associated with an extra processing step. However, their interpretations of the late positivity could be framed in the two-step model. For example, increasing attentional demands might reflect the processing of hidden negation (Clark & Chase, 1972; Grodzinsky et al., 2018; Kaup et al., 2006) or downward monotonicity (Agmon et al., 2019; Grodzinsky et al., 2018).

To summarise, the two-step model found substantial support in experimental data on processing negative sentences (Clark, 1976; Clark & Chase, 1972; Just & Carpenter, 1971; Kaup et al., 2006). However, its predictions were tested rather indirectly. To directly test the two-step model, we reanalysed EEG data from the

experiment of Augurzky et al. (2020) with the HsMM-MVPA developed by Anderson et al. (2016). This reanalysis had a two-fold goal. Firstly, we aimed to test one explanation of the late positivity found by Augurzky et al. (2020) namely, processing of negative quantifiers involves an additional step of processing related to the complexity of representation. A positive finding would support the two-step model. Secondly, we wanted to test if the procedure of truth value computation for negative quantifiers involves an additional step of processing as also predicted by the two-step model. In the next section, we elaborate on the main theoretical assumptions of HsMM-MVPA and its relation to the traditional ERP analysis.

1.3. Hidden semi-Markov model multivariate pattern analysis

The potential of HsMM-MVPA has been shown in several cognitive tasks (see Borst & Anderson, 2021, for review). In domains close to quantification, Zhang, Walsh et al. (2018) validated the HsMM-MVPA in a mathematical problem-solving task, and Berberyan et al. (2021) discovered the stages of processing in a lexical decision task. HsMM-MVPA was also applied to an associative recognition task and a Sternberg Working Memory task (Anderson et al., 2016), a perceptual speed-accuracy trade-off task (Van Maanen et al., 2021), perceptual decision-making task (Berberyan et al., 2020), working memory task (Zhang, van Vugt et al., 2018), a numerical cognition task (Groeneweg et al., 2021), and a mental rotation task (Heimisch et al., 2023). Together, the method's validity is well-established in cognitive tasks that involve only a few processing stages. The methodological advancement of the current study is to apply this method to a task in which participants have to process a stream of stimuli, such as the words of a sentence.

HsMM-MVPA has been applied in a top-down manner to test computational models of specific tasks (Anderson et al., 2016) as well as in a bottom-up manner (Walsh et al., 2017) when the computational model of the task is unknown, and the goal is to infer the number of processing stages from the neural data. In this paper, we applied a hybrid approach. On the one hand, we do not have a fully established processing model of quantified sentences that would predict a precise number of processing stages. In the absence of such a model, we applied the bottom-up strategy. On the other hand, we aimed to test a very precise, theory-driven prediction that there should be at least one more stage in one experimental condition than the other. This could be considered a top-down test of our main prediction.

HsMM-MVPA allows us to test our main hypothesis, namely the presence of an extra processing step that the two-step model predicts. This is because HsMM-MVPA aims to identify the onset of cognitive events in the EEG signal on a trial-by-trial basis, so-called *bumps* (cf., Makeig et al., 2002; Yeung et al., 2004, 2007), under the assumption that these cognitive events occur on every trial and are from the same temporal distribution. The durations between bumps (so-called *flats*) can differ under the assumption that cognitive processes vary in duration as well, e.g. because of variation in task demands between different experimental conditions. Moreover, flats are variable from trial to trial under the assumption that information processing by participants is also prone to trial-by-trial variability. Because the first bump might not occur exactly with the onset of a stimulus presentation, the first stage starts with a flat. Therefore, for n bumps, there are always $n + 1$ stages.

HsMM-MVPA can be used to estimate the most likely distribution over time of the bump locations, as well as their amplitudes. A subsequent comparison of models with different numbers of bumps provides evidence for a particular number of bumps (see the appendix in Anderson et al., 2016, for discussion and mathematical details). This entails that under the two-step hypothesis, the processing of negative quantifiers requires an additional bump, signalling an additional cognitive event.

1.3.1. HsMM-MVPA and ERP methods

Thus far, we have outlined the HsMM-MVPA at a conceptual level. To apply the method to the EEG data, Anderson et al. (2016) proposed a linking assumption between the EEG signal and bumps estimated by the HsMM-MVPA. They postulated that HsMM-MVPA identifies bumps of EEG activity, which correspond to the ERPs. This assumption is compatible with two theories of ERP generation (Makeig et al., 2002): the classical theory and the synchronised oscillation theory.

Although there is a clear theoretical relationship between HsMM-MVPA and ERP methods, the advantage of the former is that it overcomes some shortcomings of the latter. The main shortcoming of the ERP analysis is that it usually involves averaging the EEG signal (cf. Dubarry et al., 2017). Therefore, ERPs cannot account for the trial-by-trial variability in the onsets of the endogenous ERP components. As Walsh et al. (2017) argued, there are two consequences of the loss of trial-level variability due to averaging. Firstly, averaging might diminish an ERP effect if the trial-by-trial variability in the onsets of components is high. Secondly, the information about the onset of the cognitive event is lost. In

contrast to traditional ERPs, HsMM-MVPA analyses the EEG signal at the single-trial level instead of averaging it from multiple trials and participants. The HsMM-MVPA identifies on a trial-by-trial basis the bumps of EEG activity corresponding to the onsets of ERPs (cf. Anderson et al., 2016).

Needless to say, the ERP technique advanced our understanding of how language is processed. However, HsMM-MVPA is more appropriate when it comes to testing the two-step model. By comparing different HsMM-MVPA models with different numbers of bumps, we can quantify the number of processing stages in each experimental condition. No such analysis is possible with ERPs. Moreover, HsMM-MVPA detects the onsets of processing stages on the trial level, while the methods of estimating trial-by-trial ERP latencies have limitations (Walsh et al., 2017). This limitation of classical ERP can be avoided with the help of HsMM-MVPA. HsMM-MVPA makes it possible to consider the information about trial-by-trial variability in the EEG signal in the averaging by computing the bump-related potentials (BRPs, Berberyan et al., 2021).

1.3.2. Bump-related potentials

BRPs (Berberyan et al., 2021) take advantage of both HsMM-MVPA and ERP methods. Based on the above-mentioned linking assumption, the bump onsets are signatures of the ERPs. Therefore, BRPs can be computed using the trial-by-trial information about the bump onset from HsMM-MVPA. For averaging, BRPs can be computed in a fixed time window; however, in contrast to traditional ERPs, the time window is not locked to stimuli or response onsets but to the onset of the cognitive process of interest. Berberyan et al. (2021) showed that BRPs can reveal differences between experimental conditions that were lost during averaging in the traditional ERP analysis.

The purpose of BRP analysis in this study was twofold. Firstly, it allows us to link the ERP results by Augurzky et al. (2020) with the results of HsMM-MVPA and enhance the interpretation of the new findings. Secondly, by aligning the signal to the bump onsets instead of the stimuli onset we can test precisely the origin of the difference in EEG amplitude between conditions. For example, the interaction in N400 found by Augurzky et al. (2020) between the polarity of the quantifier and the truth value could be explained by the general delayed processing of the negative quantifier due to an extra step preceding this time window. Under this hypothesis, the bump link with the N400 component could out-scope the Augurzky et al. (2020) analysis time window. However, once the EEG signal is aligned to the bump onset, the difference in N400 should be visible for both quantifiers. Alternatively, if an additional processing step was not a reason for the absence of the N400 difference for *fewer than half*, we should replicate the interaction found by Augurzky et al. (2020) in the BRP analysis.

2. Methods

We applied the HsMM-MVPA to data from a picture-sentence verification task collected and analysed by Augurzky et al. (2020) to test the two-step model hypothesis directly. For a detailed description of participants, exclusion criteria, experimental design, procedure, and EEG recording, see Augurzky et al. (2020). In addition, we conducted ERP analyses to test whether we could replicate the results from Augurzky et al. (2020) study (Section 2.6) and statistical analyses of reaction time data and stage durations estimated using HsMM-MVPA (Section 2.7).

2.1. Participants

We excluded two participants due to movement EEG artifacts from the sample of 23 participants analysed by Augurzky et al. (2020).[4]

2.2. Experimental design and procedure

The experiment tested the processing of positive and negative proportional quantifiers in a picture-sentence verification task employing a within-subjects factorial design. The Quantifier (*more than half* vs. *fewer than half*) at the beginning of the sentence and the visual Context (picture A vs. picture B) were manipulated such that, given a specific picture, the sentences were true with one of the Quantifiers and false with the other. In addition, the Length (long vs. short) of the sentence was manipulated as a third factor. Out of a total of 320 trials, 160 contained short sentences and the remaining 160 long ones. The short sentences had the structure *Q of the dots are ADJ* (*Q der Punkte sind ADJ*), where *Q* was one of the two tested quantifiers and *ADJ* was an adjective referring to one of two colours shown in the picture. Pictures contained geometrical shapes (e.g. circles, triangles, rectangles) in one of two colours (e.g. red and blue, purple and orange, green and yellow) randomly paired with a container shape (e.g. semicircle, squares). The container shape was important for long sentences, which referred to shapes inside or outside it. In particular, the long sentences continued with …*that are PREP of the FORM* (…*die PREP des/der FORM sind*), where FORM referred to the container

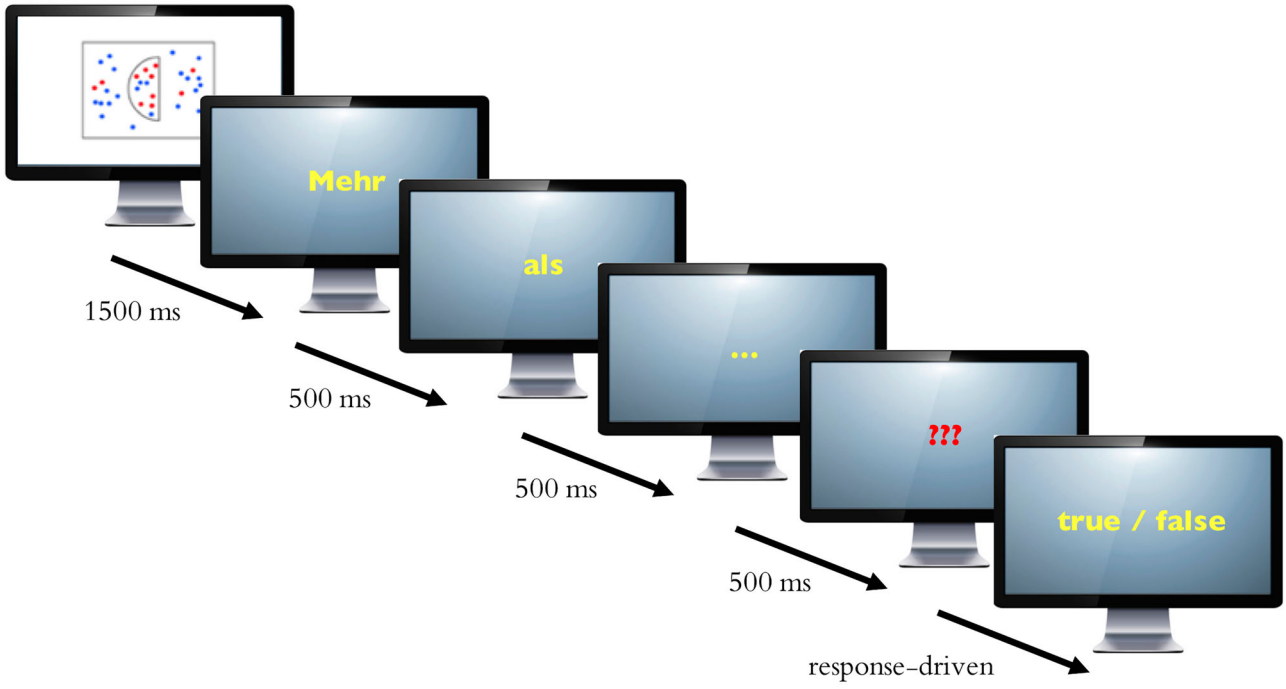

Figure 1. Procedure of the experiment (figure from Augurzky et al., 2020).

shape shown in the picture, and PREP was one of two Prepositions (inside vs. outside). The trials with short sentences thus consisted of factorial combinations of Quantifiers (*more than half*, *fewer than half*) and Truth Values (*true sentence*, *false sentence*), while trials with long sentences combined Quantifiers (*more than half*, *fewer than half*), Contexts (*picture A*, *picture B*), and Prepositions (*inside*, *outside*). Together, there were 80 trials per quantifier and per sentence length. For each combination of Quantifier x Context x Preposition, 20 pairs of context pictures were generated and paired with each sentence condition.

The experimental procedure is shown in Figure 1. Each trial started with the presentation of the context picture in the centre of the screen for 1500 ms. The sentences were presented word by word for 500 ms each and were followed by three question marks, prompting for a response. Participants evaluated the truth value of the sentences by pressing the F or J keys on a keyboard. Response-key assignment was counterbalanced across participants. They were instructed to respond as soon as possible. They did not know whether the sentences would continue until they saw the punctuation mark, which was presented on a separate screen. After participants had responded, a blank screen was displayed for 500 ms followed by three exclamation marks displayed for 1200 ms. To prevent data contamination due to eye-movement artefacts, participants were instructed to blink between trials. The experiment included a timeout procedure. The initial timeout was 1200 ms. During the experiment, the timeout was adopted to participants' response timing using exponentially weighted moving averages (Leonhard et al., 2011). Participants received feedback encouraging them to respond more quickly, i.e. the word "Faster!" (*Schneller!*) displayed on the screen, if they ran into the timeout.

2.3. Choice of analysis time windows

Previous studies (e.g. Anderson et al., 2016; Berberyan et al., 2020) have applied HsMM-MVPA from the onset of the stimuli until the response. Given that in our experiment, each word was displayed for 500 ms, it would not be possible to include whole sentences in the analysis. This would make the model too complex and the computation intractable. As mentioned in the introduction, the two-step model gives two predictions of when the extra step could occur. It could either occur during the comprehension of the sentences or during the comparison between sentences and pictures (Clark, 1976). Therefore, we chose two time windows based on previous analyses (Augurzky et al., 2020) and predictions of the two-step model.

In the first time window, we tested whether the difference in amplitude of the late positivity between *more than half* and *fewer than half* found by Augurzky et al. (2020) was related to an extra processing step during the comprehension of the quantifier. Therefore, we chose the first time window from the quantifier onset

(*more/fewer)* until 800 ms after. For this analysis, we included both short and long sentences because, at this point, the sentence type (short vs. long) did not differ physically, and participants could not have predicted the sentence type. Additionally, by including long sentences, we analysed more trials and increased the power of the analysis (typically, studies using HsMM-MVPA include at least 100 trials per condition, cf. Anderson et al., 2018). We analysed two conditions corresponding to the quantifiers *more than half* and *fewer than half*.

In the second time window, from the onset of the adjective until the response, we tested whether comparing the propositional sentence meaning and the picture representation is reflected in the processing stages. In this time window, Augurzky et al. (2020) found an interaction between quantifier polarity and sentence truth value reflected in N400 amplitude. For this analysis, we included only short sentences because, at this point, the sentence type (short vs. long) differed physically. We analysed four conditions: *more than half* true sentences, *more than half* false sentences, *fewer than half* true sentences, and *fewer than half* false sentences.

2.3.1. Expected bumps and components

As already mentioned, there are no existing computational models that predict the number of bumps during the processing of quantified sentences. However, based on previous HsMM-MVPA studies and Augurzky et al. (2020), we can make certain predictions about bumps and components that we expect to observe in the current study. Because participants performed a visual task, we expect to observe an N100 component (Luck et al., 2000). For the components related to linguistic processing, in addition to N400 and P600, we expect to observe bumps reflecting the P200 (Dambacher et al., 2006) and P300 (Jouravlev et al., 2016). We predicted that the task could elicit components related to working memory (Ruchkin et al., 1992). In addition, in the time window from the adjective onset, we expected to observe late negativity linked to the evaluation of the truth value (Wiswede et al., 2013) and response-related components such as the late positive complex (LPC) (De Jong et al., 1990) and the centroparietal positivity (CPP) (Twomey et al., 2015).

2.4. EEG data preprocessing

The data preprocessing consisted of two stages: initial data preprocessing and artifact rejection, and specific preprocessing needed for HsMM-MVPA, BRPs and ERPs. For data preprocessing, we used MATLAB R2019b and R2021a (The MathWorks, Inc.), MATLAB toolbox EEGlab 2019 and 2021 (Delorme & Makeig, 2004), and preprocessing scripts adapted from Berberyan et al. (2020).

We referenced the electrodes to mastoids. We downsampled the data from 2048 Hz to 1024 Hz and applied a 0.3 Hz high-pass filter and a 20 Hz low-pass filter. The filters and references were the same as in Augurzky et al. (2020). In the next step, we manually cleaned the data from the artifacts, except for the eye movement-related artifacts. We interpolated the signal from noisy electrodes for 8 participants. We did not interpolate more than 15% of electrodes. Following manual artifact rejection, we applied Independent Component Analysis (ICA, *runica* algorithm Delorme & Makeig, 2004; Delorme et al., 2007). We removed the components related to eye movements (usually 1 or 2 components) and components related to voltage artifacts. In this way, we removed 2 components on average.

After cleaning the data, we applied preprocessing steps specific to the HsMM-MVPA and BRPs and ERPs. For HsMM-MVPA analysis, we followed the steps from Berberyan et al. (2020). Downsampling of EEG data is a necessary preprocessing step for the HsMM-MVPA to make the computations tractable. We downsampled data to 100 Hz and removed the incorrect trials. We also removed trials with too short or too long reaction times based on the mean $+/-2$ SD criterion. Then, we applied the baseline correction of 200 ms and epoched the data.

Bump magnitudes and flats are not directly estimated from the electrode signal. Anderson et al. (2016) performed the Principal Component Analysis (PCA) to reduce the intercorrelations of the EEG signal. They included the first 10 components that accounted for the largest variance of data (above 90%). In the final step, we also performed the PCA. PCA is also used to handle the highly correlated brain signal. We included 10 first components, which accounted for 92.99% of the variance in the time window from quantifier onset and 91.07% of the variance in the time window from adjective onset. The data were normalised with a z-transformation.

For ERP and BRP analyses, we downsampled data to 100 Hz to have the same frequency as in HsMM-MVPA. Then for ERPs, we epoched the data from the stimuli onset (quantifier or adjective) until 800 ms after. For BRPs, we sequentially epoched the data from the trial-by-trial onset of each bump. To compute the average from multiple trials, we fixed the durations of the BRP analyses time window. The duration of this time window was based on the results of ERP analyses and HsMM-MVPA (see more in the Results section). We applied a baseline correction of 200 ms to ERP and BRP epoched data.

2.5. HsMM-MVPA

For the HsMM-MVPA analysis, we used MATLAB R2019b and R2021a (The MathWorks, Inc.), MATLAB toolbox EEGlab 2019 and 2021 (Delorme & Makeig, 2004), and analysis scripts adapted from Berberyan et al. (2020, 2021) and available at OSF (https://osf.io/z49me/).

Firstly, we applied HsMM-MVPA to the data from the time window from quantifier onset until 800 ms after. In this time window, the maximum number of bumps was 15.[5] We fitted the HsMM-MVPA model separately to each quantifier. The model uses the data from all participants and trials simultaneously to estimate two sets of parameters: the bump magnitudes and flat durations. Based on the 10 components that explain 90% of the variance in the data, the HsMM-MVPA model estimates 10 magnitude values for every bump. The flat durations are assumed to follow a gamma-2 distribution with a shape parameter fixed to value equals 2 and a free scale parameter. The scale parameter captures the trial-by-trial variability in flat durations. To obtain the maximum likelihood, HsMM-MVPA used the expectation–maximisation (EM) algorithm. To avoid estimation of local maxima instead of the global maximum likelihood, we applied the same procedure described by Zhang, Walsh et al. (2018) (see also Berberyan et al., 2020, 2021). Firstly, the model fitted the maximum number of bumps (n) in the time window. In the next step, the algorithm iteratively removed one bump and fitted models with bumps ($n-1$). Then, all $n-1$ bumps models were compared, and the best model was selected. The algorithm repeated this procedure until it fitted the model with only one bump.

The log-likelihood of the model increases as the complexity of the model (number of bumps) increases. To avoid overfitting, we used the leave-one-out cross-validation following the procedure of Anderson et al. (2016). The increasing complexity of the model was only justified when the more complex model fitted better to a significantly larger number of participants. This was assessed by a computing sign test on the number of participants for whom the log-likelihood of the more complex model increased. In this way, we chose a model that generalised across the largest number of participants. The sign test was used in several previous HsMM-MVPA studies (e.g. Anderson et al., 2016; Berberyan et al., 2021; Van Maanen et al., 2021). As a result of the leave-one-out cross-validation, we obtained the bump magnitudes and scale parameters of the gamma-2 distribution for each participant.

2.6. ERP and BRP statistical analyses

To test for significant effects in ERPs and BRPs, we used a cluster-based random permutation test (Maris & Oostenveld, 2007) in Fieldtrip (Oostenveld et al., 2011). The cluster-based random permutation test is a non-parametric test suitable for handling the multiple comparison problem. In the first step, for every sample (pair of channel and time point) the differences between conditions was calculated and quantified by the paired t test for dependent samples. For the analysis in the time window from quantifier onset, we compared two conditions: *fewer than half* and *more than half*. For the analysis in the time window from adjective onset, we tested interactions between Quantifier and Truth value. Therefore, firstly, we computed the main effects of Quantifier and Truth value, and in the next step, we calculated the interaction effect also using the dependent t test. In the next step, samples with a higher t value than the 0.05 thresholds were selected and clustered. The maximum cluster statistics were chosen. The Monte Carlo method was used to obtain Monte Carlo significance probability. The random partitions and computation of the cluster statistic was repeated 1000 times. The Monte Carlo p value was calculated and compared to the conventional p value at the level 0.05.

Following the findings of Augurzky et al. (2020), we selected four regions of interest (ROIs): left anterior (ROI 1: F3, F7, FC1, FC5), right anterior (ROI 2: F4, F8, FC2, FC6), left posterior (ROI 3: CP1, CP5, C3, P3), and right posterior (ROI 4: CP2, CP6, C4, P4).

In the ERP analysis in the time window from the quantifier onset, we expected to find a significant difference in positivity between *fewer than half* and *more than half* around 450 to 800 ms from quantifier onset. In the analysis in the time window from the adjective onset, we expected to find a significant interaction between Quantifiers and Truth value in the 300 to 400 ms time window from the adjective onset. We predicted that this difference would be reflected in the N400 for the *more than half* false sentence condition compared to the *more than half* true sentence condition, and a lack of difference for *fewer than half*.

Moreover, we tested whether the differences of interest mentioned above will be replicated in the BRP analyses. The goal of these analyses was to show that these bumps can be linked with the cognitive process underlying the difference found in ERPs. Therefore, we selected one bump for the quantifier onset time window that preceded the expected late positivity effect and one for the adjective onset time window that preceded the expected N400 effect.

2.7. Statistical analysis of reaction times and stage durations

The main goal of our analysis was to test the prediction that sentences with *fewer than half* have at least one more stage of processing than sentences with *more than half*. We expected the verification of sentences with a negative quantifier to take longer than with a positive quantifier because of the extra processing step.

Therefore, we tested the differences in reaction time data. We selected the same trials as for the HsMM-MVPA and for ERP analysis. We ran a linear mixed-effects model with Quantifier (*more than half*, *fewer than half*), Truth value (*true*, *false*) and Quantifier x Truth value interaction as predictors, and tested their effects on the reaction time data. We included the by-subject random intercepts, and tested the significance of the random slope for the trial (centred). To interpret the main effects, we used contrast coding.

Moreover, we aimed to link stage durations with reaction times. We expected that some stage durations might be related to the specific cognitive processes that affect the length of reaction times, while other stages could just reflect the fixed processing pattern of the upcoming input (such as encoding and motor preparation). While the latter stages are not particularly meaningful for our hypothesis, the former could give us insight into differences in quantifier processing.

Using a linear mixed-effects regression model, we tested whether the stage durations (in the time window from adjective onset) predicted the reaction times for each experimental condition. We applied a backward fitting procedure. First, we included all stages as predictors. Then we excluded the insignificant predictors one by one according to their p values (we excluded the predictors with higher p values first) until only significant predictors were left in the model.

Finally, we selected those stages that were significant predictors of reaction times in every experimental condition and tested the differences in their duration between conditions: *more than half* true sentences, *more than half* false sentences, *fewer than half* true sentences, and *fewer than half* false sentences. We ran mixed-effects models on each stage with the predictors (contrast coded): Quantifier (*more than half*, *fewer than half*), Truth value (*true*, *false*) and their interaction. We also included the by-subject random intercept and the by-subject random slope for the trial (centred) if it was significant.

For all analyses described in this section, we used the *lmer* function from the R package *lmerTest* (Kuznetsova et al., 2017) to run a regression model and *anova* function for model comparison. We log-transformed reaction time data[6] and the stage distributions.

3. Results

3.1. Preliminary analyses

The goal of preliminary analyses was to demonstrate the polarity effect in behavioural and EEG data. Moreover, we aimed to replicate the ERP findings of Augurzky et al. (2020) in 100 Hz frequency to compare the results with HsMM-MVPA.

3.1.1. Reaction time analysis

We found a significant main effect of Quantifier ($\beta = -0.15$, $t = -5.33$, $p < 0.001$), a significant interaction ($\beta = -0.15$, $t = -2.64$, $p = 0.008$),[7] and a significant intercept ($\beta = 5.79$, $t = 48.50$, $p < 0.001$). The effect of Truth value was not significant ($\beta = -0.05$, $t = -1.83$, $p = 0.07$). The verification of *fewer than half* was slower than the verification of *more than half* (see Figure 2). Moreover, the effect of Truth value went in the opposite direction for two quantifiers: reaction times were slower for false responses in *more than half* and faster in *fewer than half*.

3.1.2. ERP analyses

Our ERP analysis from quantifier onset replicated the finding of Augurzky et al. (2020). We also found a greater late positivity for *fewer than half* than *more than half* between 450 and 800 ms after stimulus onset (see Figure 2 in Appendix 2.1). We observed this effect in all regions of interest. However, the difference was more prominent on the centro-parietal electrodes.

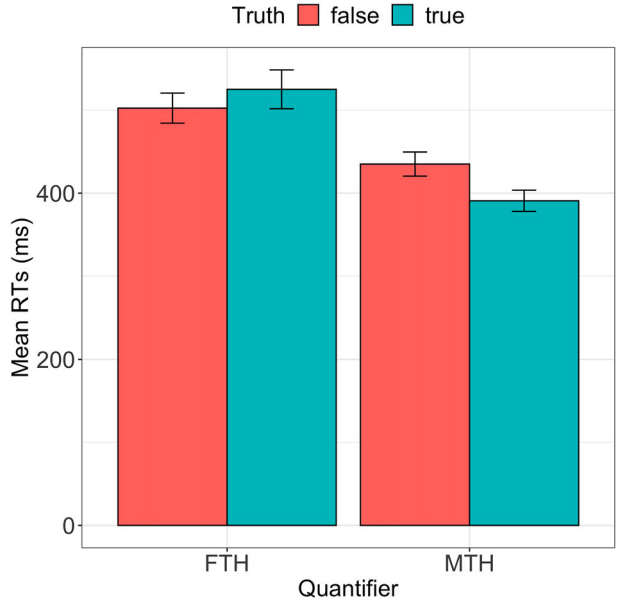

Figure 2. Mean reaction times for short sentences (*fewer than half* is abbreviated as FTH and *more than half* as MTH). The error bars represent within-participant SE.

In addition to the replicated late positivity, we also found a difference in EEG amplitude around 200 ms from quantifier onset at some electrodes. The amplitude was higher for *fewer than half* than for *more than half*, which could reflect the difference in P200 potential between quantifiers (see Figure 2 in Appendix 2.1).

In ERP analyses in the time window post adjective onset, we found an interaction effect between Quantifier and Truth value between 300 to 400 ms in three regions of interest (the interaction was insignificant only in ROI 1, see Figure 4 in Appendix 2.2). This finding shows a greater negative potential for *more than half* false sentences compared to *more than half* true sentences. Our non-parametric analysis, therefore, replicated the N400 effect found by Augurzky et al. (2020). In addition, we found a main effect of Truth value in all ROIs between 300 and 400 ms after the adjective onset.

Moreover, we found an interaction effect in ROIs 3 and 4 in the later time window between 450 and 800 ms from stimuli onset. The EEG amplitude was lower for *more than half* true sentences compared to *more than half* false sentences. In addition, the main effect of Quantifier in the same time window in ROI 1 and the effect of Truth value in ROI 2 were significant.

Finally, we also found an interaction effect around 200 ms from the adjective onset in the ROI 4 at one electrode. This effect was not reported previously by Augurzky et al. (2020).

3.1.3. The summary of preliminary analyses

To summarise, we found evidence for the polarity effect in both preliminary analyses. The polarity effect in behavioural data manifested in the longer reaction times for *fewer than half* than for *more than half*. Moreover, the ERP analyses revealed the polarity effect and interaction between the polarity of a quantifier and the truth value of a sentence in the EEG data. In the time window from quantifier onset, we replicated the late positivity effect and also showed the P200 effect. In the time window from the adjective onset, we found an interaction effect between the sentence Truth value and Quantifier between 300 to 400 ms and between 450 and 800 ms. We replicated, thus, the Augurzky et al. (2020) findings in a lower sampling frequency and using a different statistical analysis that controls well for the false alarm rate and limits the changes for the false positive result (Maris & Oostenveld, 2007). Together, the results encourage the stages of processing analysis.

3.2. Test of the two-step model with HsMM-MVPA

3.2.1. Quantifier onset

We fitted the HsMM-MVPA to two conditions *more than half* (average of 118 trials per subject) and *fewer than half* (average of 103 trials per subject) separately. The leave-one-out cross-validation analysis revealed that for *more than half*, the model with 8 bumps (9 stages) had the highest mean log-likelihood (LL $= -213.244$). This model had improved fit for a significant number of participants (17 out of 21 participants, sign test $p<0.05$) compared to the model with 7 bumps. For *fewer than half*, the results were not unequivocal. The model with 8 bumps (8 bumps LL $= -193.672$) had an improved fit for only 11 out of 21 participants and did not significantly outperform the model with 7 bumps (7 bumps LL $= -193.696$) as indicated by a sign test (sign test $p>0.05$). Importantly, there was no evidence in favour of the 9-bumps model because this model did not have a better fit over the 8-bumps model for any participant (LL $= -1432.21$). Because the modelling solution for *fewer than half* was ambiguous between 7 and 8 bumps, we ran an additional analysis in which we fitted one HsMM-MVPA model to the combined data from both quantifiers (average of 221 trials per subject). We found that the 8-bump model was better than the 7-bump model for 18 out of 21 participants (sign test $p<0.05$), and it had the highest mean log-likelihood (LL $= -394.585$) out of all models. It also outperformed the separate models (sum of mean LL $= -406.916$). Therefore, we are inclined to select a model with 8 bumps for both quantifiers. Together, this finding does not support the hypothesis that *fewer than half* has more processing stages than *more than half*.

Figure 3 presents the topologies and average stage duration for the 8-bump (9-stage) model fitted to quantifiers separately. The figure shows that the processing time courses and bump topologies of the two quantifiers are comparable. Figure 4 shows the topologies and average stage duration for the 8-bump (9-stage) combined model.

3.2.2. Adjective onset

We followed the same model comparison procedure as in the first time window. Firstly, we fitted the HsMM-MVPA to four conditions: *more than half* true sentences (average of 29 trials per subject), *more than half* false sentences (average of 29 trials per subject), *fewer than half* true sentences (average of 23 trials per subject), and *fewer than half* false sentences (average of 28 trials per subject) separately. For all conditions, the model with 10 bumps (11 stages) had the highest log-likelihood: *fewer than half* false LL $= -77.0794$ (better model for

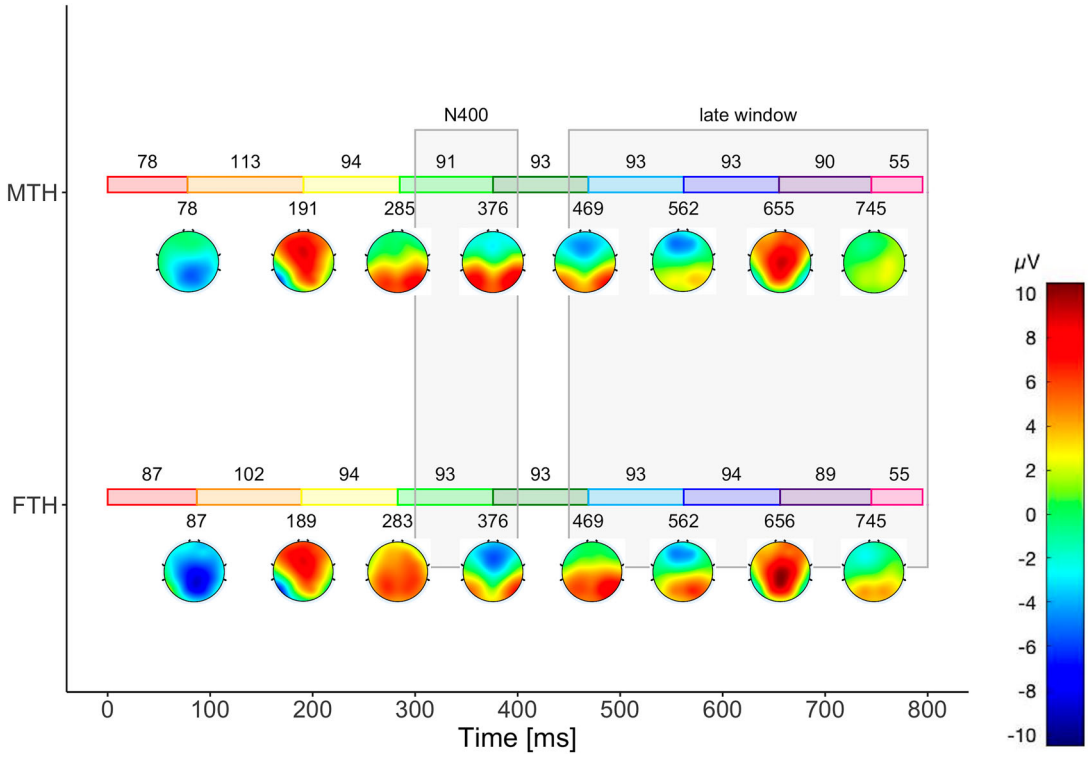

Figure 3. Bump topologies and stage durations for separate models *more than half* (MTH) and *fewer than half* (FTH) from quantifier onset. The values above bump topologies correspond to the average onset of the bump. The coloured bars indicate the stage durations. The values above the coloured bars show the mean stage durations. Additionally, the gray lines indicate the ERP analysis time windows from Augurzky et al. (2020).

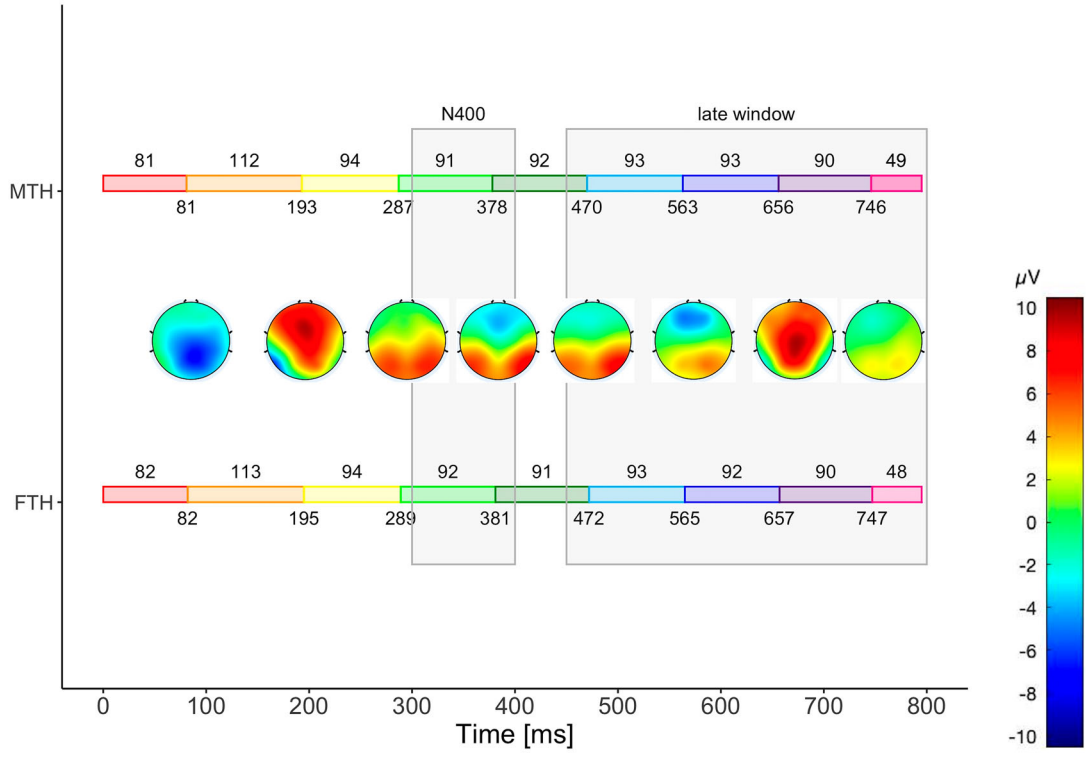

Figure 4. Bump topologies and stage durations for the combined model from quantifier onset plotted in each condition separately, *more than half* (MTH) and *fewer than half* (FTH). The coloured bars indicate the stage durations. The values above the coloured bars show the mean stage durations, and the values below the average onset of the bumps. Additionally, the gray lines indicate the ERP analysis time windows from Augurzky et al. (2020).

13 participants than the simpler model), *fewer than half* true LL $= -72.4346$ (better model for 14 participants than the simpler model), *more than half* false LL $= -75.4606$ (better model for 13 participants than the simpler model), and *more than half* true LL $= -40.9431$ (better model for 14 participants than the simpler model). However, there was a substantial variability in model fit between participants. The models with the highest mean log-likelihood were not better for a significant number of participants (sign test $p>0.05$).

In the next step, we fitted a combined model to all four conditions together (average of 110 trials per subject). Because of the great variability in model fit, we wanted to test whether the 10-bump model would fit all conditions equally well. The 10-bump model with mean log-likelihood of LL $= -246.924$ fitted the data best for a significant number of participants (19 out of 21, sign test $p<0.05$). The 10-bump combined model from adjective onset was better for 15 out of 21 participants than the separate models (sum of mean LL $= -265.9177$), meaning that the more complex, separate models did not outperform the combined, simpler model.

We plotted the bump topologies and stage durations of the separate 10-bump models in Figure 5 and of the combined model in Figure 6. We observed that the variability in stages durations is greater for separate models than for the combined model, however, in both cases the durations of the first 6 processing stages as well as Stages 8 and 10 were rather fixed across conditions. Stages 7 and 11 seemed to vary across conditions in both separate models and the combined model. The bumps topologies for *fewer than half* in separate models were very similar (Figure 5). Moreover, the topologies of bumps for *more than half* false sentence were also similar to those of *fewer than half* with the exception of the fourth bump, which was more negative in *more than half* false sentence condition. This negative bump was absent in *more than half* true sentence condition. Figure 6 shows that the topologies of bumps replicated in the combined model, however, because the combined model assumes that all bumps are shared across conditions, the contrast in the topology of fourth bump disappeared.

3.2.3. HsMM-MVPA summary

The HsMM-MVPA in both time windows did not support the predictions derived from the two-step model. We did not find evidence for an extra processing step for *fewer than half* compared to *more than half*. In both time windows, the more parsimonious combined model outperformed the more complex, separate models. Despite the variability in separate model fits, we did not find support for more processing stages for *fewer than half* than *more than half*. Still, in the time

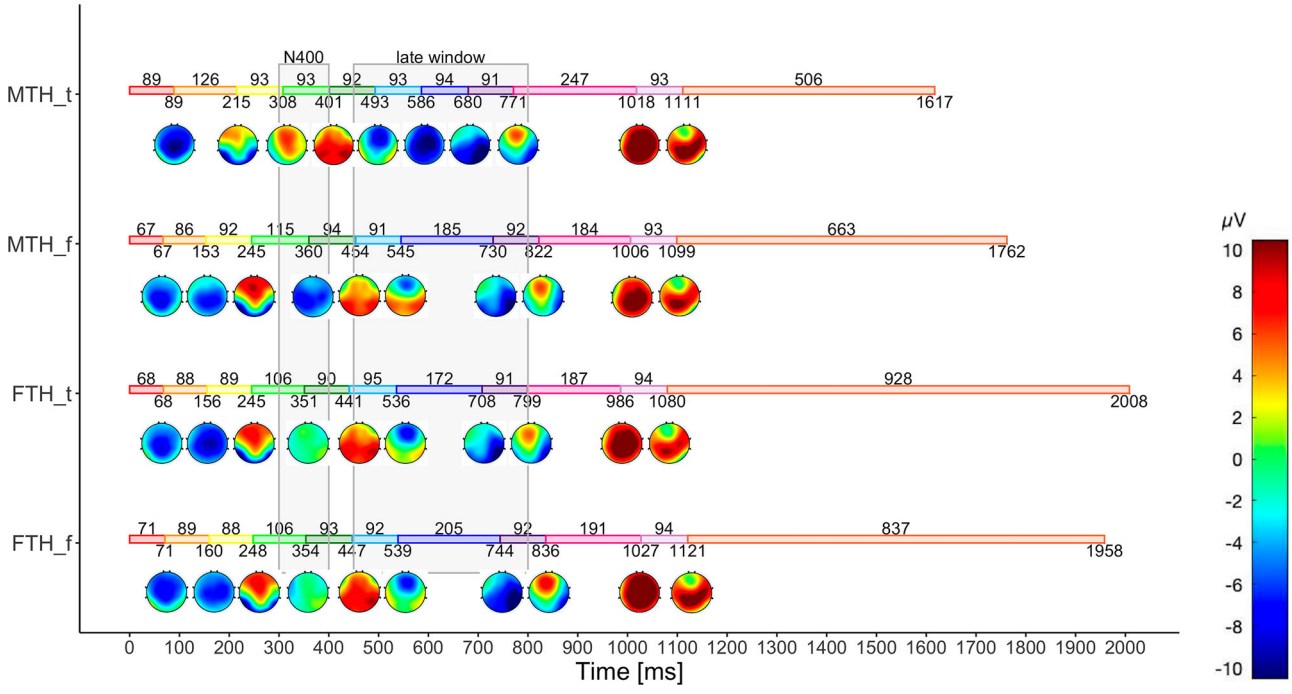

Figure 5. Bump topologies and stage durations for separate models *more than half* true sentence (MTH_t), *more than half* false sentence (MTH_f), *fewer than half* true sentence (FTH_t), and *fewer than half* false sentence (FTH_f) from adjective onset until the response. The values above bump topologies correspond to the average onset of the bump. The coloured bars indicate the stage durations. The values above the coloured bars show the mean stage durations. Additionally, the gray lines indicate the ERP analysis time windows from Augurzky et al. (2020).

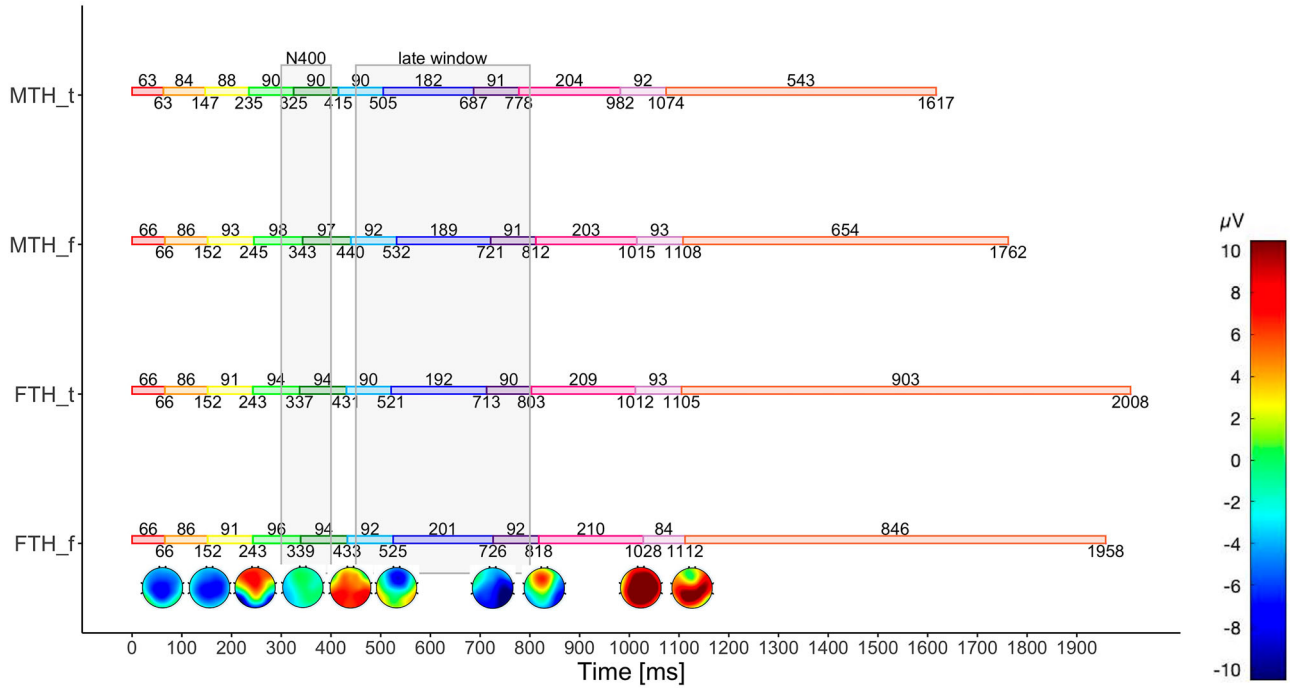

Figure 6. Bump topologies and stage durations for the combined model from adjective onset until the response plotted separately for each condition, *more than half* true sentence (MTH_t), *more than half* false sentence (MTH_f), *fewer than half* true sentence (FTH_t), and *fewer than half* false sentence (FTH_f). The timing is relative to adjective onset. The coloured bars indicate the stage durations. The values above the coloured bars show the mean stage durations, and the values below the average onset of the bumps. Additionally, the grey lines indicate the ERP analysis time windows from Augurzky et al. (2020).

window from the quantifier onset, we even found more evidence for fewer processing stages for *fewer than half* than *more than half*. Nonetheless, our reaction time analysis (see Section 3.1.1) indicated that *fewer than half* was verified more slowly than *more than half*. We hypothesised that this difference should be reflected in the duration of processing stages (see Section 3.5). Moreover, our ERP analyses demonstrated the polarity effect in EEG data (see Section 3.1.2) To link the ERP findings with the HsMM-MVPA results, we performed BRP analyses.

3.3. BRP analyses

3.3.1. Quantifier onset

In the quantifier time window, the difference in EEG amplitude between *more than half* and *fewer than half* was detected between 450 to 800 ms from the quantifier onset. Therefore, for the BRP analysis, we selected Bump 4, which preceded this time window (the average onset of this bump was 378 ms for *more than half* and 381 ms for *fewer than half* based on the combined model, see Figure 4). We aligned the EEG signal to trial-by-trial onsets of this bump and chose a time window of 400 ms duration from the bump onset to cover a similar time interval as in the ERP analysis.

Our BRP analysis replicated the effects found in ERP analysis in all ROIs (see Figure 3 in Appendix 2.1). The EEG amplitude was higher for *fewer than half* than for *more than half*.

3.3.2. Adjective onset

In the adjective time window, the difference in N400 between *more than half* and *fewer than half* was detected between 300 to 400 ms from the adjective onset. Therefore, for the BRP analysis, we selected Bump 3, which preceded the N400 time window analysis. We aligned the EEG signal to trial-by-trial onsets of this bump based on the parameters estimated from the combined model. The average Bump 3 onsets were 243 ms for *fewer than half*, 243 ms for *more than half* false, and 235 ms for *more than half* true (see Figure 6). We chose a time window from the bump onset until 560 ms after to cover a similar time interval as in the ERP analysis.

Similarly to the ERP analysis, the BRPs revealed an interaction between sentence truth value and quantifier, as well as late negativity for *more than half* true sentences (see Figure 5 in Appendix 2.2).

3.3.3. BRP summary

Our BRP analyses replicated the findings of the ERP analyses. The first BRP analysis showed that in the absence

of a difference in the number of stages or their duration, the processing of *fewer than half* was associated with a larger late positivity than the processing of *more than half*. Moreover, the second BRP analysis confirmed that the N400 effect was present only for *more than half* false sentences. This suggests, in contrast to two-step model predictions, the N400 was absent in the case of *fewer than half* sentences, not delayed. Overall, these findings show the compatibility of the ERP analyses and HsMM-MVPA and speak against the two-step model prediction.

3.4. Do stages predict the length of reaction times?

Based on the HsMM-MVPA results, we concluded that the difference in reaction times between conditions can not be attributed to an additional processing step. In the subsequent analysis, we tested whether the polarity effect observed as the reaction time differences between quantifiers (see Section 3.1.1) can be explained by the difference in duration of some cognitive stages. This analysis was exploratory. Because the best-fitting model was the combined model, we used the stage durations extracted from this model (see Figure 6) as predictors.

3.4.1. Fewer than half *false*

The final model[8] included Stage 4 ($\beta = 0.35$, $t = 2.33$, $p = 0.02$), Stage 7 ($\beta = 0.29$, $t = 3.90$, $p = 0.0001$), Stage 9 ($\beta = 0.40$, $t = 5.34$, $p < 0.0001$), and Stage 11 ($\beta = 0.34$, $t = 14.64$, $p < 0.0001$) as significant predictors of reaction times. The intercept of the model was not significant ($\beta = -1.42$, $t = -1.80$, $p = 0.07$). The reaction times for *fewer than half* false sentences were thus predicted by Stages 4, 7, 9, and 11.

3.4.2. Fewer than half *true*

The final model[9] included Stage 6 ($\beta = 0.89$, $t = 4.97$, $p < 0.0001$), Stage 7 ($\beta = 0.15$, $t = 2.20$, $p = 0.03$), Stage 9 ($\beta = 0.39$, $t = 6.94$, $p < 0.0001$), and Stage 11 ($\beta = 0.35$, $t = 18.69$, $p < 0.0001$) as significant predictors of reaction times. The intercept of the model was also significant ($\beta = -3.03$, $t = -3.91$, $p = 0.0001$). The reaction times for *fewer than half* true sentences were thus predicted by Stages 6, 7, 9, and 11.

3.4.3. More than half *false*

The final model[10] included Stage 6 ($\beta = 0.51$, $t = 3.20$, $p = 0.001$), Stage 7 ($\beta = 0.43$, $t = 6.81$, $p < 0.0001$), Stage 9 ($\beta = 0.33$, $t = 5.84$, $p < 0.0001$), Stage 10 ($\beta = 0.29$, $t = 1.99$, $p = 0.047$), and Stage 11 ($\beta = 0.32$, $t = 17.10$, $p < 0.0001$) as significant predictors of reaction times. The intercept of the model was also significant ($\beta = -3.58$, $t = -3.90$, $p = 0.0001$). The reaction times for *more than half* false sentences were thus predicted by Stages 6, 7, 9, 10, and 11.

3.4.4. More than half *true*

The final model[11] included Stage 2 ($\beta = 0.49$, $t = 2.85$, $p = 0.004$), Stage 4 ($\beta = 0.46$, $t = 3.09$, $p = 0.002$), Stage 7 ($\beta = 0.18$, $t = 2.53$, $p = 0.01$), Stage 9 ($\beta = 0.56$, $t = 9.02$, $p < 0.0001$), and Stage 11 ($\beta = 0.390$, $t = 16.63$, $p < 0.0001$) as significant predictors of reaction times. The intercept of the model was also significant ($\beta = -4.60$, $t = -4.96$, $p < 0.0001$). The reaction times for *more than half* true sentences were thus predicted by Stages 2, 4, 7, 9, and 11.

3.5. Stage durations analysis

In the final step of the analysis, we tested whether a significant main effect of Quantifier and an interaction between Quantifier and Truth value found in the reaction times (see Section 3.1.1) would be reflected in the duration of stages. We tested stages extracted from the combined model that were significant predictors of reaction times for all experimental conditions, namely Stages 7, 9, and 11 (see also Figure 6).

3.5.1. Stage 7

Firstly, we tested the differences in Stage 7.[12] The interaction between Quantifier and Truth value was not significant ($\beta = 0.003$, $t = 0.11$, $p = 0.91$, $\chi^2(1) = 0.01$, $p = 0.91$). The best model had a significant intercept ($\beta = 5.18$, $t = 191.86$, $p < 0.001$) and two main effects of Quantifier ($\beta = -0.03$, $t = -2.14$, $p = 0.03$) and Truth value ($\beta = -0.04$, $t = -2.71$, $p = 0.007$). We found that *fewer than half* had longer Stage 7 than *more than half* and that this stage was longer for false sentences compared to true sentences (see Figure 7).

3.5.2. Stage 9

Secondly, we tested Stage 9.[13] We found that the Quantifier x Truth value interaction was not significant ($\beta = 0.006$, $t = 0.21$, $p = 0.84$, $\chi^2(1) = 0.04$, $p = 0.84$). The model without interaction had a significant intercept ($\beta = 5.25$, $t = 176.55$, $p < 0.001$), but neither a main effect of Quantifier ($\beta = -0.02$, $t = -1.51$, $p = 0.13$), nor a main effect of Truth value ($\beta = -0.01$, $t = -0.77$, $p = 0.44$). We conclude that this stage did not differ between experimental conditions (see Figure 7).

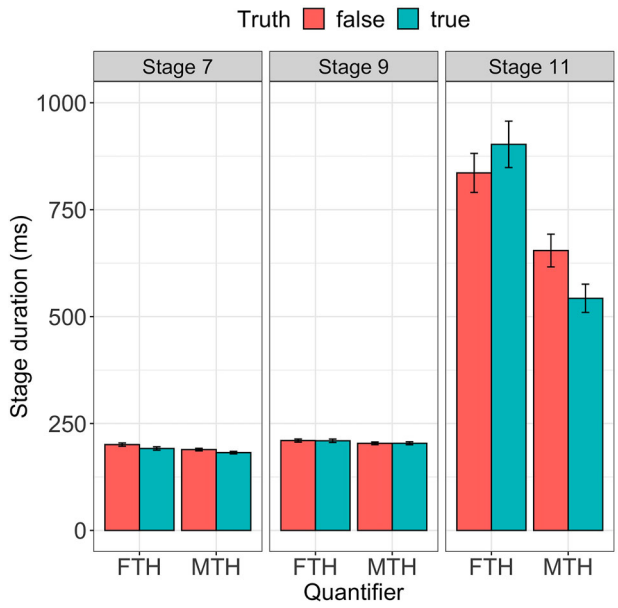

Figure 7. Mean durations of stages 7, 9, and 11. *Fewer than half* is abbreviated as FTH and *more than half* as MTH. The error bars represent within-participant SE.

3.5.3. Stage 11

Next, we tested Stage 11.[14,15] We found that the Quantifier x Truth value interaction was not significant ($\beta = -0.10$, $t = -1.24$, $p = 0.22$, $\chi^2(1) = 1.52$, $p = 0.22$). The model without interaction had a significant intercept ($\beta = 5.80$, $t = 42.56$, $p < 0.001$), and a main effect of Quantifier ($\beta = -0.23$, $t = -5.47$, $p < 0.001$) and insignificant main effect of Truth value ($\beta = -0.04$, $t = -1.05$, $p = 0.30$). We found that *fewer than half* had a longer Stage 11 than *more than half* (see Figure 7).

3.5.4. Combined effect of stages 7, 9, and 11

Finally, we summed the durations of Stages 7, 9, and 11 to test whether their combined duration could be predicted by the effect of Quantifier and Truth value. This effectively tests whether the reaction time effect can be completely attributed to those stages.[16] The interaction effect was not significant ($\chi^2(1) = 1.36$, $p = 0.24$). After excluding the interaction from the model, we found that the main effect of Quantifier ($\beta = -0.15$, $p < 0.001$) and intercept ($\beta = 6.70$, $p < 0.001$) were significant. Still, the main effect of Truth value was not ($\beta = -0.04$, $p = 0.10$).

3.5.5. The summary of stage durations analyses

The analyses of stage durations revealed that Stages 7, 9, and 11 extracted from the combined model explained some variability found in the reaction times. Moreover, Stages 7 and 11 were longer for *fewer than half* than *more than half*, showing that the polarity effect was present at the level of specific processing stages.

4. Discussion

The main goal of the current study was to test the two-step model prediction that the polarity effect, namely difficulties related to the processing of negative vs. positive quantifiers, is due to an extra processing step that negative quantifiers involve. Such an extra step could be related to a higher complexity of the representation of negative quantifiers due to the hidden negation or to a longer verification procedure (e.g. Clark, 1976; Grodzinsky et al., 2018). To test this hypothesis, we analysed data from a picture-sentence verification task collected by Augurzky et al. (2020). We used a novel method, the HsMM-MVPA (Anderson et al., 2016), to detect the processing stages in the EEG signal and directly compared these stages between quantifiers. Our analysis demonstrated the polarity effect but did not support the two-step model. In the next sections of the discussion, we summarise our main findings and give a functional interpretation of the processing stages that contributed to the explanation of the polarity effect (see functional interpretation of all processing stages in Appendix 4). Moreover, we propose an explanation of our results in the light of an alternative account that attributes the source of the polarity effect to contextual factors. Finally, we elaborate on the methodological implications of our study.

4.1. Test of the two-step model

Firstly, we analysed the time window from quantifier onset until 800 ms after. We fitted separate HsMMs for each quantifier. Our finding in this time window did not support the hypothesis that there is an extra processing step in the comprehension of *fewer than half*. Both separate models favoured the 8-bump solution, although the superiority of the 8-bump solution was weaker for the negative quantifier, and the model with 7 bumps was equally good. Crucially, we did not find support for the 9-bump solution for *fewer than half*. Secondly, we found that the 8-bump model with combined conditions of both quantifiers fitted the data better than the separate models. In addition, we observed that the onsets of bumps were very similar across conditions and that there was little variation in stage durations between quantifiers (see Figures 3 and 4). Therefore, we endorse the conclusion that the time course of processing the quantifier at the beginning of the sentence is

fixed, and the processing of *fewer than half* was not delayed.

In addition to the quantifier position, we also analysed the time window from the adjective onset until the response. The 10-bump model had the highest mean log-likelihood for all quantifier and truth value combination conditions. We also fitted a combined model to all conditions jointly and established that the model with 10 bumps fitted the data most accurately. The combined model also outperformed the four separate models, meaning that, in the time window from the adjective onset, we again did not find support for an extra processing step for the negative quantifier.

4.2. Differences in processing of positive and negative quantifiers

The results of the HsMM-MVPA did not support the two-step model. However, our additional analyses showed some differences in the processing of negative and positive quantifiers. The polarity effect was present in reaction time data and EEG data. In this section, we discuss the evidence for the polarity effect and the functional interpretation of the stages associated with this effect (see the functional interpretation of all processing stages in Appendix 4).

In the ERP and BRP analyses, we replicated the previous result of Augurzky et al. (2020). There was a greater late positivity for *fewer than half* than *more than half*. The process underlying the observed late positivity had already started with the onset of Bump 4. Importantly, the onset of this bump was comparable between *fewer than half* and *more than half*, suggesting that the processing of *fewer than half* was not delayed at this point. The late positivity continued in Stage 6, which began with a parietal distributed positive bump. The increasing positive activity may reflect the P600 potential often found in sentence-level linguistic tasks (see Brouwer et al., 2017, 2012, for review). In our study, the positive parietal activity is characteristic for four consecutive stages (Stages 4 to 7). This observation is consistent with literature that shows multiple functional interpretations of the P600 and is probably linked to different underlying components (Regel et al., 2014). For example, the P600 is sensitive to ungrammatical structures as well as pragmatic manipulation e.g. presupposition triggers (Domaneschi et al., 2018; Jouravlev et al., 2016), or irony (Regel et al., 2014). The so-called pragmatic P600 is preceded by the P200 component (also cf. Jouravlev et al., 2016 the P3b/P600 complex). Moreover, while the syntactic P600 is widespread over the scalp, the pragmatic P600 is mostly visible on central and parietal electrodes. However, both components have a similar latency of around 500 ms after the stimuli onset. We found a pattern of results characteristic to the pragmatic P600: a large peak of the P200 component with a significant difference between quantifiers in Stage 3 (see the ERP analysis in Figure 2 in the Appendix, results of the HsMM-MVPA in Figure 4 for stages onsets, and Appendix 4), and a P600 difference in the time window of Stage 6. We therefore suggest that Stage 6 can be interpreted as a stage of processing pragmatic properties (we will come back to this point in Section 4.3).

Moreover, in the time window from the adjective onset, we replicated the ERP effects found by Augurzky et al. (2020). We found the N400 effect only for *more than half* false sentences and later amplitude deflection for *more than half* true sentences. The HsMM-MVPA showed that Stages 4 and 5 jointly contributed to the N400 effect captured in the ERP analysis. The BRP analysis replicated the N400 effect when the EEG signal was aligned with the trial-by-trial onset of Stage 4. The topology of Bump 4 in the *more than half* false sentence condition resembles the N400 component (see Figure 5). The N400 was captured in ERP analysis (Appendix, Figure 4) and, crucially, in BRP analysis (Appendix, Figure 5) as well. Based on this analysis, we arrived at the interpretation of the fifth stage as the semantic encoding stage.

In addition, we demonstrated that the polarity effect was present in reaction time data. Verifying *fewer than half* took longer than the verification of *more than half*. We further explored the differences between conditions in the stage durations estimated with HsMM-MVPA. We found that Stages 7, 9, and 11 predicted reaction times across conditions.

We found the polarity effect in Stage 7. We replicated the larger negativity for *more than half* true sentences found by Augurzky et al. (2020) in the time window from the adjective onset that overlaps with Stage 7. Moreover, this stage was longer for *fewer than half* than for *more than half* and for false than for true sentences. Therefore, we suggest that this stage is related to a cognitive process that is crucial to the truth evaluation. For example, this stage could reflect a comparison between a picture and a sentence in working memory. As mentioned in Section 2.3.1, we expected to observe negative components related to the truth evaluation of the sentence. The Bump 6 with a negative frontal distribution could reflect the working memory processes (Ruchkin et al., 1992). Participants involved their working memory in the comparison between picture and sentence representations. Our findings align with the recent hypothesis that negative quantifiers add

more load to working memory than positive ones (Agmon et al., 2022).

As mentioned in Section 2.3.1, we expected to observe bumps related to response preparation and execution. Stages 9 and 11 could be related to these processes because they also contribute to the explanation of the polarity effect in reaction times. Stage 9 begins with a left frontal positive bump, which may reflect the expected late positive complex (LPC De Jong et al., 1990). The LPC is related, among others, to the successful inhibition of a response (Kiefer et al., 1998). The response inhibition in the current experiment might be related to the fact that participants had to wait for the three question marks signal to respond.

The last stage, Stage 11, ends with the participants' response. The differences in duration of this stage between conditions reflected the Quantifier effect but not the Truth value and interaction effects. Previous HsMM-MVPA studies (Anderson et al., 2016; Berberyan et al., 2021) showed that the response stage is short. In contrast, this stage was the longest in our analysis. The atypical duration of the response stage in our experiment could be explained by the higher complexity of the task or the specific experimental procedure employed. Participants may have made a decision before the onset of the signal to respond, and they drifted their attention elsewhere (cf., Kaup et al., 2006).

To sum up, our results contribute to understanding the time course of the processing of quantified sentences. Our analyses clearly demonstrated behavioural and EEG evidence for the polarity effect. At the same time, our analyses indicate that this effect is not due to an additional processing step that negative quantifiers involve. Therefore, we suggest an alternative interpretation of this effect in the light of a pragmatic account.

4.3. Alternative explanations of the polarity effect

As in the case of the two-step model, we use the label "pragmatic account" to indicate a broader class of proposals that explain the differences between negatives and affirmatives in terms of the speaker's pragmatic preferences and interaction between processed sentence and discourse context. For example, the expectation-based account (e.g. Nieuwland, 2016) proposes that negation is usually unexpected and thus generates the processing cost. The pragmatic account does not predict the number of processing stages testable with HsMM-MVPA as the two-step model does. Therefore, we only highlight the compatibility of our results with this theoretical proposal.

The pragmatic account found support in EEG studies on negation and quantifiers. Pragmatic information influences the processing of negation more than the processing of affirmative sentences (Orenes et al., 2016). For example, Nieuwland and Kuperberg (2008) showed in an EEG experiment that difficulties in processing negation disappear in a pragmatically licensed context. Moreover, negative quantifiers (such as *few*) can also be processed fully incrementally in an appropriate discourse context (Urbach et al., 2015). Nieuwland (2016) demonstrated that the difficulties of processing quantifiers are dependent on the predictability of the sentence continuation.

Augurzky et al. (2020) argued that the complexity of the quantifiers mediated the participants' ability to build expectations about the sentence. In other words, participants were more efficient in formulating predictions about *more than half* sentences than *fewer than half*. Because of the higher complexity of *fewer than half*, the generation of expectations was delayed.

Our findings substantially extend the expectation-based interpretation of Augurzky et al. (2020). As mentioned in the introduction, Augurzky et al. (2020) provided two possible explanations of how predictions could have been generated. According to one of these explanations compatible with our findings, participants encoded the picture in terms of the greater set (Clark, 1976). Based on the picture encoding, they could have immediately generated the expectations for *more than half* sentences, but not for *fewer than half* sentences. The late positivity can reflect the attempt to build expectations for *fewer than half* when it becomes clear that a negative quantifier has to be processed. The extra effort leads to the engagement of more cognitive resources and differences in EEG amplitude. This attempt is not fully successful, as reflected by the lack of N400 difference on the adjective for *fewer than half* sentences.

This explanation is consistent with our findings. Participants' expectations about sentence continuations were reflected in an N400 effect observed in the time window from adjective onset (ERP and BRP analyses). However, neither in HsMM-MVPA nor in BRP analysis did we observe a processing delay for *fewer than half*. We hypothesise thus that the lack of the N400 effect for *fewer than half* false sentences (ERP and BRP analyses) could indicate that the generation of expectations may not have been successful. Our findings give further insights into the interpretation of the N400 in Augurzky et al. (2020) study. While some studies link the N400 effect with the truth evaluation process (Augurzky et al., 2017, 2020), our results are in line with the access/retrieval account for N400 (Brouwer et al., 2017, 2012). According to this account, N400 reflects lexical retrieval but not the integration process (Delogu et al.,

2019), and it is not directly modulated by the truth value of the sentence (Kounios and Holcomb, 1992).

Finally, the larger negativity for *more than half* true sentences observed in Stage 7 after the adjective onset and the differences in duration of this stage between conditions can also be explained by referring to expectations. To successfully perform that verification task, participants had to retain in memory the information about the quantifier, the numerosity of objects, and the adjective. The expectations and default encoding of the picture in terms of a larger proportion influence the working memory load (cf. Clark & Chase, 1972). The verification of *more than half* true sentences requires retaining only the matching information in working memory (expected quantifiers, larger set, and the colour of the larger set). To verify *fewer than half*, participants had to retain information about the smaller set as well. The additional working memory load explains the quantifier effect in the duration of Stage 7 (cf. Agmon et al., 2022 similar explanation referring to working memory load). To verify false sentences, participants had to carry additional information that the colour of the set they encoded and the adjective colour did not match. This explains the effect of the truth value.[17]

While there is general agreement that successful language comprehension requires building predictions about the upcoming linguistic input (Grisoni et al., 2017, 2021), the explanation of mechanisms behind the expectations generation is less understood. The dynamic pragmatic account by Tian et al. (2010) proposes such a mechanism by referring to the so-called Questions Under Discussion (QUDs) (Roberts, 2012). This account assumes that language users process the information that has already been accommodated into the discourse context by the relevant QUDs faster. According to this approach, positive questions are considered by comprehenders as the default, because they are more frequent than negative questions. For example, when verifying the sentence "The glass is not empty," comprehenders assume that the relevant QUD is "Is the glass empty?" rather than "Is the glass not empty?". As mentioned earlier, the P600 effect observed in Stage 6 after the quantifier onset could be related to the processing of pragmatic properties of negative quantifies. One of these properties could be the accommodation of the non-default QUD triggered by the negative expression. To directly test the dynamic pragmatic account, we would have to introduce an explicit manipulation of the QUDs that would affect the encoding of the picture. The exploration of how the manipulation of QUDs would affect the stage durations can be tested in future work.

4.4. Methodological implications

Our study has several methodological implications. Firstly, it showed the dissociation between behavioural measures (reaction times) and EEG-based measures (stage durations). Secondly, we demonstrated the benefit of using different EEG data analysis methods. For example, we observed that differences in EEG signal amplitude between quantifiers (in ERP and BRP analyses) were not due to an additional processing step and were not reflected in differences in stage durations.

The HsMM-MVPA was previously applied to tasks such as associative recognition (Anderson et al., 2016; Van Maanen et al., 2021), lexical decision (Berberyan et al., 2021), or perceptual decision (Berberyan et al., 2020) tasks that involved a few cognitive stages. In contrast, we applied the HsMM-MVPA to a sentence-level task. This posed additional challenges to our analysis. For example, we could not analyse the processing of the whole sentence, but we had to select constrained time windows. Moreover, the processing stages related to one stimulus overlapped with the display of new stimuli.

In the next subsections, we elaborate on these two key methodological aspects in more detail. We also discuss the limitations of our study.

4.4.1. The dissociation between different measures

We found different effects in the reaction time data and the duration of the stages. In the behavioural data, we found an interaction between Quantifier and Truth value (see also Section 4.4.2). Next, we tried to explain this effect with the stage durations estimated from HsMM-MVPA (see Section 3.5). Three stages contributed to explaining reaction times across conditions. However, none of the stages alone nor their combined duration reflected the interaction effect found in reaction time data. This might have happened because the interaction effect in reaction times could be an effect of a unique combination of multiple stages, different for each condition, which were not included in the regression model. This finding indicates that the comparison between mean reaction times can be misleading because reaction times are affected by many different processing stages. Previous studies showed that the mapping between processing stage durations and differences in reaction times is complex. For example, Zhang et al. (2017) found in an associative recognition task that one experimental manipulation can have an opposite effect on two processing stages. While the effect of the manipulation is obscured in reaction times measure, it can be observed in HsMM-MVPA.[18] In

sum, these findings suggest that the combined effect of multiple stages can explain reaction times.

Moreover, we showed the complexity of mapping the psychological concepts such as processing difficulties on different measures of brain activity such as EEG signal amplitude, number of processing stages, or durations of these stages. In the time window from quantifier onset, we replicated the results of Augurzky et al. (2020). We found a greater late positivity in the ERP analysis for *fewer than half* than for *more than half*. Moreover, in the BRP analysis, in which the EEG signal was aligned to the trial-by-trial onset of Bump 4, we also showed that processing of *fewer than half* was associated with a greater late positivity than *more than half*. Importantly, the greater neural activity for *fewer than half* was not reflected in stage duration differences between quantifiers.

The lesson to be learned from our analyses is that the mapping between the length of particular cognitive processes, reaction times, and the amplitude of the EEG signal might be equivocal. To better understand the relationship between different measures, we suggest jointly analysing the behavioural and neural data in future studies.

4.4.2. Methodological limitations

We notice several methodological limitations of our study. Firstly, for some of the HsMM-MVPA models, we had a relatively small number of trials per condition compared to previous studies (e.g. Anderson et al., 2016; Berberyan et al., 2020; Van Maanen et al., 2021).[19] We acknowledge that in the time window from the adjective onset, the number of trials that were input to the separate models was substantially smaller than in the typical HsMM-MVPA study. However, in the time window from quantifier onset, the separate models were fit to the typical number of trials (on average, for each subject, there were more than 100 trials per condition), yet the extra step of processing was not detected. This means that while the results of separate models in the time window from adjective onset should be treated with caution, the results in the time window from quantifiers onset are fully robust.

Secondly, we noticed the greatest variability in model fit was present in the time window from adjective onset for separate models. The variability in model fit of separate models could be explained from methodological and theoretical perspectives. Methodologically, the problem with model fit could result from an insufficient number of trials per condition. Theoretically, the variation in model fit across participants could reflect individual differences in the processing of quantified sentences. Individual differences were found previously in semantic representations of vague quantifiers (Ramotowska et al., 2023) and in the processing of pragmatically underinformative sentences (Spychalska et al., 2016). Given the small number of trials per condition as the input to these HsMM-MVPA models, we can not exclude any of the discussed interpretations.

In addition, the variation in the data was also present in reaction times (see Appendix 1). The timeouts changed the typical reaction time distribution, and the log-transformation failed to fully compensate for it (e.g. the normality of residuals assumption was not fully satisfied). When the reaction times that exceeded the timeout were excluded from the analysis, the interaction effect between Quantifier and Truth value was not significant (see Appendix 1). Moreover, the analysis of Stage 11 with excluded timeout reaction times did not reveal differences between experimental conditions. This suggests that the long responses at least partially drove the differences in reaction times and Stage 11.

Participants' response strategies may have also introduced a greater variability into the EEG data. One source of variability may be movement artifacts, even though we cleaned the data and removed trials containing larger artifacts. The variability could also be a result of the experimental procedure. For example, while participants waited for a signal to respond, their attention may have drifted away from the task. Moreover, the timeout procedure may have pressured participants' responses. Augurzky et al. (2020) argued against the speed-accuracy trade-off. Nonetheless, the deadlines for the response affect the decision-making process, as shown by previous studies where participants were asked to make decisions under time pressure (Katsimpokis et al., 2020; Miletić & Van Maanen, 2019).

The timeout trials constituted a large proportion of trials (see Appendix 1). Moreover, we noticed that they were more frequent for *fewer than half* than for *more than half*. The tendency for participants to exceed the time limit more for one quantifier could be a result of processing difficulties associated with it. This poses a challenge to the interpretation of the reaction times that exceeded the time limit. On the one hand, if the timeout reaction times are due to participants' attention drifting away before responding, it would be better to exclude trials with a timeout. The focus of attention can lead to different processing stages (Van Maanen et al., 2021). However, we would have to exclude the timeout trials from all of our analyses because we do not know at what timepoint participants were distracted. This would lead to a significant reduction in the number of trials. On the other hand, the timeout

reaction times were associated more with *fewer than half* than *more than half*. They could also be a relevant source of information about processing differences. From a methodological perspective, the decision to exclude timeout trials is equivocal.

Finally, while our findings speak against an extra processing step for negative quantifiers, they can not rule out the two-step model hypothesis for other types of negatives. In particular, the two-step model originated from the studies on sentential negation. For example, Agmon et al. (2022) found that the polarity effect decreases (but does not fully diminish) for sentential negation when participants are given more time to process a sentence. No such effect was found for quantificational negation, meaning that the additional time did not aid the processing. It would be desirable to test for an extra processing step when negation refers explicitly to the to-be-negated state of affairs. Once explicitly mentioned, the representation of the to-be-negated state of affairs could be activated and processed in the extra stage. In conclusion, further studies should test the two-step model predictions in different experimental setups and with various types of negatives.

4.5. Conclusions

According to the two-step processing hypothesis, the hidden negation in negative quantifiers increases their complexity and gives rise to the polarity effect. In this study, we challenged this hypothesis in quantified sentences. By using a novel method to analyse the EEG data, we estimated the number of processing stages for sentences with the two quantifiers *more than half* and *fewer than half*. To the best of our knowledge, this is the first study that directly tested the difference in processing stages in quantified sentences and directly addressed the two-step processing hypothesis. We provided an interpretation of the processing stages and linked them to the predictions of the expectation-based account.

Notes

1. Note that both operators *-er* and and *little* are downward entailing operators: *-er* is a comparative downward entailing operator, and *little* is a negation operator.
2. See similar predictions in Szymanik and Zajenkowski (2013).
3. See similar predictions in Barwise and Cooper (1981).
4. For details about ethics approvals and informed consent see Augurzky et al. (2020).
5. We noted that 800 ms divided into 50 ms should result in a maximum number of bumps 16, not 15. The 15-bump maximal model is a result of the downsampling. The shortest time window had in some trials 79 samples, not 80. Therefore, 79 samples divided into 5 samples gave a 15 bump maximal model.
6. During the data analysis, we observed that the reaction times had somewhat bimodal distributions with large proportions of long reaction times that were not fully excluded with the outliers procedure. We ran a separate mixed-effects model on reaction times, which were not classified as timeouts. See Appendix 1 for details of this analysis.
7. In the mixed-effects regression model of RT, we included by-subject random slope for trial as it significantly improve model fit ($\chi^2(1) = 255.49$; $p < 0.001$).
8. For *fewer than half* false sentences, we did not include the by-subject random slope for trial ($\chi^2(1) = 0.73$, $p = 0.39$) because it did not significantly improve model fit.
9. For *fewer than half* true sentences, we included the by-subject random slope for trial ($\chi^2(1) = 20.20$, $p < 0.001$).
10. For *more than half* false sentences, we included the by-subject random slope for trial ($\chi^2(1) = 10.73$, $p = 0.001$).
11. For *more than half* true sentences, we included the by-subject random slope for trial ($\chi^2(1) = 11.08$, $p = 0.001$).
12. We included the by-subject random slope for trial ($\chi^2(1) = 22.49$, $p < 0.001$).
13. We included the by-subject random slope for trial ($\chi^2(1) = 25.92$, $p < 0.001$).
14. We observed the same problem with Stage 11 distribution as with reaction time distribution. The stage of the distribution was binomial even after log-transformation, and, therefore, the model did not fit the assumption of normal distribution of residuals. We ran an additional analysis on the data without timeout reaction times in Appendix 3
15. We included the by-subject random slope for trial ($\chi^2(1) = 92.14$, $p < 0.001$).
16. We found that the random slope for trial was significant ($\chi^2(1) = 126.08$, $p < 0.001$).
17. Additional support for our interpretation of Stage 7 comes from the analysis of long sentences (cf. Ramotowska, 2022).
18. See also Berberyan et al. (2020) and Van Maanen et al. (2021) for comparison between processing stages estimated using HsMM-MVPA and another computational model.
19. See also the discussion about the sufficient number of trials to estimate parameters of other cognitive models, e.g. Wagenmakers (2009), Lerche et al. (2017), Osth et al. (2017) and Boehm et al. (2018).

Acknowledgments

Conceptualisation – SR, PA, FS, LvM, JS; Data curation – SR, FS; Formal analysis – SR, KA, HB; Funding acquisition -- PA, JS; Investigation – PA, SR; Methodology – SR, PA, FS, HB, LvM, JS; Software – SR, KA, HB; Supervision – LvM, JS; Visualisation – SR, HB; Writing – original draft – SR; Writing – review & editing – SR, KA, PA, FS, HB, LvM, JS.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

This research received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement n. STG 716230 CoSaQ and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 75650358 – SFB 833, project B1.

Data availability statement

The data are available at https://talar.sfb833.uni-tuebingen.de/erdora/cmdi/SFB833/B01/polarity_files.Analysis scripts are available at https://osf.io/6amgu/?view_only=d766908a62ef4f4cbe3c532adb5b7af2.

ORCID

S. Ramotowska http://orcid.org/0000-0003-3381-4089
F. Schlotterbeck http://orcid.org/0000-0002-2310-9151

References

Agmon, G., Loewenstein, Y., & Grodzinsky, Y. (2019). Measuring the cognitive cost of downward monotonicity by controlling for negative polarity. *Glossa: A Journal of General Linguistics*, *4*(1), 36. https://doi.org/10.5334/gjgl.770.

Agmon, G., Loewenstein, Y., & Grodzinsky, Y. (2022). Negative sentences exhibit a sustained effect in delayed verification tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(1), 122–141.

Anderson, J. R., Borst, J. P., Fincham, J. M., Ghuman, A. S., Tenison, C., & Zhang, Q. (2018). The common time course of memory processes revealed. *Psychological Science*, *29*(9), 1463–1474. https://doi.org/10.1177/0956797618774526

Anderson, J. R., Zhang, Q., Borst, J. P., & Walsh, M. M. (2016). The discovery of processing stages: Extension of Sternberg's method. *Psychological Review*, *123*(5), 481–509. https://doi.org/10.1037/rev0000030

Augurzky, P., Bott, O., Sternefeld, W., & Ulrich, R. (2017). Are all the triangles blue?-ERP evidence for the incremental processing of German quantifier restriction. *Language and Cognition*, *9*(4), 603–636. https://doi.org/10.1017/langcog.2016.30

Augurzky, P., Schlotterbeck, F., & Ulrich, R. (2020). Most (but not all) quantifiers are interpreted immediately in visual context. *Language, Cognition and Neuroscience*, *35*(9), 1203–1222. https://doi.org/10.1080/23273798.2020.1722846

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, *4*(2), 159–219. https://doi.org/10.1007/BF00350139

Beltrán, D., Morera, Y., García-Marco, E., & De Vega, M. (2019). Brain inhibitory mechanisms are involved in the processing of sentential negation, regardless of its content. Evidence from EEG theta and beta rhythms. *Frontiers in Psychology*, *10*, 1782. https://doi.org/10.3389/fpsyg.2019.01782

Berberyan, H. S., Van Maanen, L., van Rijn, H., & Borst, J. (2020). EEG-based identification of evidence accumulation stages in decision-making. *Journal of Cognitive Neuroscience*, *33*(3), 510–527. https://doi.org/10.1162/jocn_a_01663

Berberyan, H. S., van Rijn, H., & Borst, J. P. (2021). Discovering the brain stages of lexical decision: Behavioral effects originate from a single neural decision process. *Brain and Cognition*, *153*, 105786. https://doi.org/10.1016/j.bandc.2021.105786

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., & Wagenmakers, E. J. (2018). Estimating across-trial variability parameters of the diffusion decision model: Expert advice and recommendations. *Journal of Mathematical Psychology*, *87*, 46–75. https://doi.org/10.1016/j.jmp.2018.09.004

Borst, J. P., & Anderson, J. R. (2021). Discovering cognitive stages in M/EEG data to inform cognitive models. In B. Forstmann & B. Turner (Eds.), *An introduction to model-based cognitive neuroscience* (2nd ed.,). Springer.

Botvinick, M. M., Carter, C. S., Braver, T. S., Barch, D. M., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, *41*(S6), 1318–1352. https://doi.org/10.1111/cogs.2017.41.issue-S6

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

Clark, H. H. (1976). *Semantics and comprehension*. Mouton & Co.B.V.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*(3), 472–517. https://doi.org/10.1016/0010-0285(72)90019-9

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. https://doi.org/10.1016/j.brainres.2006.02.010

De Jong, R., Coles, M. G., Logan, G. D., & Gratton, G. (1990). In search of the point of no return: The control of response processes. *Journal of Experimental Psychology. Human Perception and Performance*, *16*(1), 164–182. https://doi.org/10.1037/0096-1523.16.1.164

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

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, *34*, 1443–1449. https://doi.org/10.1016/j.neuroimage.2006.11.004

Deschamps, I., Agmon, G., Loewenstein, Y., & Grodzinsky, Y. (2015). The processing of polar quantifiers, and numerosity perception. *Cognition*, *143*, 115–128. https://doi.org/10.1016/j.cognition.2015.06.006

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

Journal of Neurolinguistics, *45*, 13–35. https://doi.org/10.1016/j.jneuroling.2017.08.002

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*(C), 412–431. https://doi.org/10.1016/0001-6918(69)90065-1

Dubarry, A. S., Llorens, A., Trebuchon, A., Carron, R., Liégeois-Chauvel, C., Bénar, C., & Alario, F. X. (2017). Estimating parallel processing in a language task using single-trial intracerebral electroencephalography. *Psychological Science*, *4*(28), 414–426. https://doi.org/10.1177/0956797616681296

Dudschig, C., & Kaup, B. (2018). How does "not left" become "right"? Electrophysiological evidence for a dynamic conflict-bound negation processing account. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(5), 716–728.

Farshchi, S., Andersson, A., van de Weijer, J., & Paradis, C. (2020). Processing sentences with sentential and prefixal negation: An event-related potential study. *Language, Cognition and Neuroscience*, *36*(1), 84–98.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*(4), 400–409. https://doi.org/10.1111/psyp.1983.20.issue-4

Grisoni, L., Mccormick Miller, T., & Pulvermüller, F. (2017). Neural correlates of semantic prediction and resolution in sentence processing. *The Journal of Neuroscience*, *37*(18), 4848–4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017

Grisoni, L., Tomasello, R., & Pulvermüller, F. (2021). Correlated brain indexes of semantic prediction and prediction error: Brain localization and category specificity. *Cerebral Cortex*, *31*(3), 1553–1568. https://doi.org/10.1093/cercor/bhaa308

Grodzinsky, Y., Agmon, G., Snir, K., Deschamps, I., & Loewenstein, Y. (2018). The processing cost of Downward Entailingness: The representation and verification of comparative constructions. *ZAS Papers in Linguistics*, *60*, 435–451. https://doi.org/10.21248/zaspil.60.2018.475

Grodzinsky, Y., Jaichenco, V., Deschamps, I., Sánchez, M. E., Fuchs, M., Pieperhoff, P., & Amunts, K. (2020). Negation and the brain. In V. Dèprez & T. M. Espinal (Eds.), *The Oxford handbook of negation* (pp. 693–712). Oxford University Press.

Groeneweg, E., Archambeau, K., & Van Maanen, L. (2021). A hidden semi-Markov model classifier for strategy detection in multiplication problem solving. In *Proceedings of the International Conference on Cognitive Modeling* (pp. 302–308).

Hackl, M. (2009). On the grammar and processing of proportional quantifiers: Most versus more than half. *Natural Language Semantics*, *17*(1), 63–98. https://doi.org/10.1007/s11050-008-9039-x

Hagoort, P., Brown, C., & 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

Heim, I. (2006). Little. In M. Gibson & J. Howell (Eds.), *In Proceedings of SALT 16*. CLC Publications.

Heimisch, L., Preuss, K., & Russwinkel, N. (2023). Cognitive processing stages in mental rotation—How can cognitive modelling inform HsMM-EEG models?*Neuropsychologia*, *188*, 108615. https://doi.org/10.1016/j.neuropsychologia.2023.108615

Jouravlev, O., Stearns, L., Bergen, L., Eddy, M., Gibson, E., & Fedorenko, E. (2016). Processing temporal presuppositions: An eventrelated potential study. *Language, Cognition and Neuroscience*, *31*(10), 1245–1256. https://doi.org/10.1080/23273798.2016.1209531

Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 244–253. https://doi.org/10.1016/S0022-5371(71)80051-8

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

Katsimpokis, D., Hawkins, G. E., & Van Maanen, L. (2020). Not all speed-accuracy trade-off manipulations have the same psychological effect. *Computational Brain and Behavior*, *3*(3), 252–268. https://doi.org/10.1007/s42113-020-00074-y

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, *38*(7), 1033–1050. https://doi.org/10.1016/j.pragma.2005.09.012

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2007). The experiential view of language comprehension: How is negation represented? In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (pp. 255–288). Lawrence Erlbaum Associates.

Kiefer, M., Marzinzik, F., Weisbrod, M., Scherg, M., & Spitzer, M. (1998). The time course of brain activations during response inhibition: Evidence from event-related potentials in a go/no go task. *Neuroreport*, *9*(4), 765–770. https://doi.org/10.1097/00001756-199803090-00037

Kounios, J., & Holcomb, P. J. (1992). Structure and process in semantic memory: Evidence from event-related brain potentials and reaction times. *Journal of Experimental Psychology: General*, *121*(4), 459–479. https://doi.org/10.1037/0096-3445.121.4.459

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*(1), 621–647. https://doi.org/10.1146/psych.2011.62.issue-1

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Leonhard, T., Fernández, S. R., Ulrich, R., & Miller, J. (2011). Dual-task processing when task 1 is hard and task 2 is easy: Reversed central processing order? *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 115–136.

Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods*, *49*(2), 513–537. https://doi.org/10.3758/s13428-016-0740-2

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive*

Sciences, *4*(11), 432–440. https://doi.org/10.1016/S1364-6613(00)01545-X

Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T. (2002). Dynamic brain sources of visual evoked responses. *Science*, *295*(5555), 690–694. https://doi.org/10.1126/science.1066168

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Miletić, S., & Van Maanen, L. (2019). Caution in decision-making under time pressure is mediated by timing ability. *Cognitive Psychology*, *110*, 16–29. https://doi.org/10.1016/j.cogpsych.2019.01.002

Nieuwland, M. S. (2016). Quantification, prediction, and the online impact of sentence truth-value: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(2), 316–334.

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, *19*(12), 1213–1218. https://doi.org/10.1111/j.1467-9280.2008.02226.x

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Orenes, I., Moxey, L., Scheepers, C., & Santamaría, C. (2016). Negation in context: Evidence from the visual world paradigm. *Quarterly Journal of Experimental Psychology*, *69*(6), 1082–1092. https://doi.org/10.1080/17470218.2015.1063675

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

Osth, A. F., Dennis, S., & Heathcote, A. (2017). Likelihood ratio sequential sampling models of recognition memory. *Cognitive Psychology*, *92*, 101–126. https://doi.org/10.1016/j.cogpsych.2016.11.007

Palaz, B., Rhodes, R., & Hestvik, A. (2020). Informative use of "not" is N400-blind. *Psychophysiology*, *57*(12), e13676. https://doi.org/10.1111/psyp.v57.12.

Ramotowska, S. (2022). *Quantifying quantifier representations: Experimental studies, computational modeling, and individual differences* [Doctoral dissertation]. https://eprints.illc.uva.nl/id/eprint/2202/1/DS-2022-03.text.pdf.

Ramotowska, S., Steinert-Threlkeld, S., van Maanen, L., & Szymanik, J. (2023). Uncovering the structure of semantic representations using a computational model of decision-making. *Cognitive Science*, *47*(1), e13234. https://doi.org/10.1111/cogs.v47.1

Regel, S., Meyer, L., & T. C. Gunter (2014). Distinguishing neurocognitive processes reflected by P600 effects: Evidence from ERPs and neural oscillations. *PloS One*, *9*(5), e96840. https://doi.org/10.1371/journal.pone.0096840

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics*, *5*, 1–69.

Ruchkin, D. S., Johnson, R., Grafman, J., Canoune, H., & Ritter, W. (1992). Distinctions and similarities among working memory processes: An event-related potential study. *Cognitive Brain Research*, *1*(1), 53–66. https://doi.org/10.1016/0926-6410(92)90005-C

Schlotterbeck, F., Ramotowska, S., van Maanen, L., & Szymanik, J. (2020). Representational complexity and pragmatics cause the monotonicity effect. In S. Denison, M. Mack, Y. Xu, & B. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 3398–3404). Cognitive Science Society.

Spychalska, M., Kontinen, J., & Werning, M. (2016). Investigating scalar implicatures in a truth-value judgement task: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, *31*(6), 817–840. https://doi.org/10.1080/23273798.2016.1161806

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*(C), 276–315. https://doi.org/10.1016/0001-6918(69)90055-9

Szymanik, J., & Zajenkowski, M. (2013). Monotonicity has only a relative effect on the complexity of quantifier verification. In M. Aloni, M. Franke, & F. Roelofsen (Eds.), *Proceedings of the 19th Amsterdam Colloquium* (pp. 219–225). University of Amsterdam.

Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why we simulate negated information: A dynamic pragmatic account. *Quarterly Journal of Experimental Psychology*, *63*(12), 2305–2312. https://doi.org/10.1080/17470218.2010.525712

Tucker, D., Tomaszewicz, B., & Wellwood, A. (2018). Decomposition and processing of negative adjectival comparatives. In E. Castroviejo, L. McNally, & G. Weidman Sassoon (Eds.), *The Semantics of Gradability, Vagueness, and Scale Structure: Experimental Perspectives* (Vol. 4, pp. 243–273). Springer.

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643. https://doi.org/10.1111/ejn.2015.42.issue-1

Urbach, T. P., DeLong, K. A., & Kutas, M. (2015). Quantifiers are incrementally interpreted in context, more than less. *Journal of Memory and Language*, *83*, 79–96. https://doi.org/10.1016/j.jml.2015.03.010

Urbach, T. P., & Kutas, M. (2010). Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, *63*(2), 158–179. https://doi.org/10.1016/j.jml.2010.03.008

Van Maanen, L., Portoles, O., & Borst, J. P. (2021). The discovery and interpretation of evidence accumulation stages. *Computational Brain & Behavior*, *2021*, 1–21.

Van Maanen, L., & Van Rijn, H. (2010). The locus of the Gratton effect in picture–word interference. *Topics in Cognitive Science*, *2*(1), 168–180. https://doi.org/10.1111/tops.2010.2.issue-1

Van Maanen, L., van Rijn, H., & Taatgen, N. (2012). RACE/A: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive Science*, *36*(1), 62–101. https://doi.org/10.1111/cogs.2012.36.issue-1

Wagenmakers, E. J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology*, *21*(5), 641–671. https://doi.org/10.1080/09541440802205067

Walsh, M. M., Gunzelmann, G., & J. R. Anderson (2017). Relationship of P3b single-trial latencies and response times in one, two, and three-stimulus oddball tasks. *Biological Psychology*, *123*, 47–61. https://doi.org/10.1016/j.biopsycho.2016.11.011

Wason, P. C. (1961). Response to affirmative and negative binary statements. *British Journal of Psychology*, *52*(2), 133–142. https://doi.org/10.1111/bjop.1961.52.issue-2

Wiswede, D., Koranyi, N., Müller, F., Langner, O., & Rothermund, K. (2013). Validating the truth of propositions: Behavioral and ERP indicators of truth evaluation processes. *Social Cognitive and Affective Neuroscience*, *8*(6), 647–653. https://doi.org/10.1093/scan/nss042

Yeung, N., Bogacz, R., Holroyd, C. B., & Cohen, J. D. (2004). Detection of synchronized oscillations in the electroencephalogram: An evaluation of methods. *Psychophysiology*, *41*(6), 822–832. https://doi.org/10.1111/psyp.2004.41.issue-6

Yeung, N., Bogacz, R., Holroyd, C. B., Nieuwenhuis, S., & Cohen, J. D. (2007). Theta phase resetting and the error-related negativity. *Psychophysiology*, *44*(1), 39–49. https://doi.org/10.1111/psyp.2007.44.issue-1

Young, R., & Chase, W. (1971a). *Additive stages in the comparison of sentences and pictures*. Paper presented at Midwestern Psychological Association meetings.

Young, R., & Chase, W. G. (1971b). Additive stages in the comparison of sentences and pictures. In *Midwestern Psychological Association Meetings*. Chicago.

Zhang, Q., van Vugt, M., Borst, J. P., & Anderson, J. R. (2018). Mapping working memory retrieval in space and in time: A combined electroencephalography and electrocorticography approach. *NeuroImage*, *174*, 472–484. https://doi.org/10.1016/j.neuroimage.2018.03.039

Zhang, Q., Walsh, M. M., & Anderson, J. R. (2017). The effects of probe similarity on retrieval and comparison processes in associative recognition. *Journal of Cognitive Neuroscience*, *29*(2), 352–367. https://doi.org/10.1162/jocn_a_01059

Zhang, Q., Walsh, M. M., & Anderson, J. R. (2018). The impact of inserting an additional mental process. *Computational Brain & Behavior*, *1*(1), 22–35. https://doi.org/10.1007/s42113-018-0002-8

Zylberberg, A., Dehaene, S., Roelfsema, P. R., & Sigman, M. (2011). The human Turing machine: A neural framework for mental programs. *Trends in Cognitive Sciences*, *15*(7), 293–300.